My research over Bayesian Optimization and Gaussian Processes

Eduardo C. Garrido–Merchán
PhD student. Teacher Assistant. Universidad Autónoma de Madrid

Index

- Introduction: Bayesian Optimization and Gaussian Processes.

Index

- Introduction: Bayesian Optimization and Gaussian Processes.
- Predictive Entropy Search for Constrained Multiobjective Scenarios.

Index

- Introduction: Bayesian Optimization and Gaussian Processes.
- Predictive Entropy Search for Constrained Multiobjective Scenarios.
- Dealing with Integer and Categorical Valued Variables in GP-BO.

Index

- Introduction: Bayesian Optimization and Gaussian Processes.
- Predictive Entropy Search for Constrained Multiobjective Scenarios.
- Dealing with Integer and Categorical Valued Variables in GP-BO.
- Parallel PESMOC.

Index

- Introduction: Bayesian Optimization and Gaussian Processes.
- Predictive Entropy Search for Constrained Multiobjective Scenarios.
- Dealing with Integer and Categorical Valued Variables in GP-BO.
- Parallel PESMOC.
- Applications over Bayesian Networks, Wave Energy and Cooking.

Index

- Introduction: Bayesian Optimization and Gaussian Processes.
- Predictive Entropy Search for Constrained Multiobjective Scenarios.
- Dealing with Integer and Categorical Valued Variables in GP-BO.
- Parallel PESMOC.
- Applications over Bayesian Networks, Wave Energy and Cooking.
- Ideas for DarkMachines unsupervised learning project.

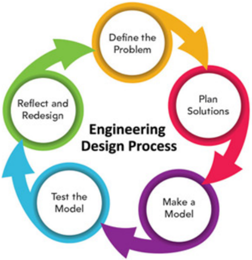
Challenges in Engineering Design

The society demands new products of better quality, functionality, usability, etc.!

Challenges in Engineering Design

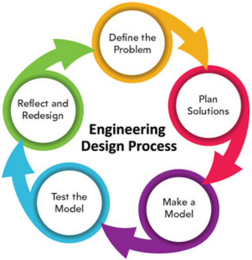
The society demands new products of better quality, functionality, usability, etc.!



Challenges in Engineering Design

The society demands new products of better quality, functionality, usability, etc.!



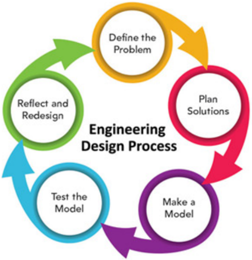



- Many choices at each step.

Challenges in Engineering Design

The society demands new products of better quality, functionality, usability, etc.!

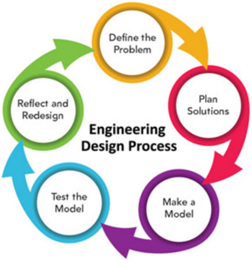


- Many choices at each step.
- Complicated and high dimensional.

Challenges in Engineering Design

The society demands new products of better quality, functionality, usability, etc.!

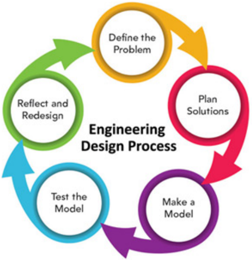


- Many choices at each step.
- Complicated and high dimensional.
- Difficult for individuals to reason about.

Challenges in Engineering Design

The society demands new products of better quality, functionality, usability, etc.!

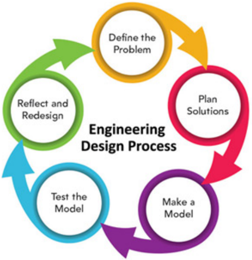


- Many choices at each step.
- Complicated and high dimensional.
- Difficult for individuals to reason about.
- Prone to human bias.

Challenges in Engineering Design

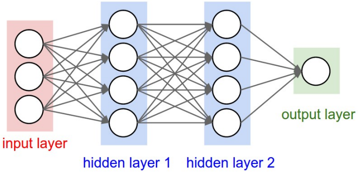
The society demands new products of better quality, functionality, usability, etc.!

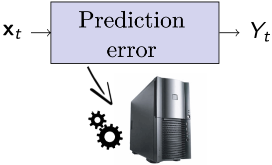


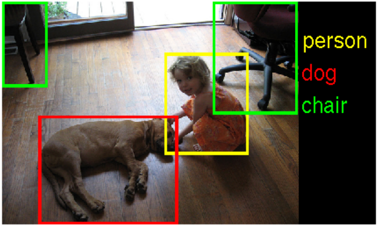
- Many choices at each step.
- Complicated and high dimensional.
- Difficult for individuals to reason about.
- Prone to human bias.

Optimization is a challenging task in new products design!

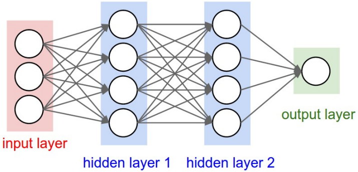
Example: **Deep Neural Network** for object recognition.

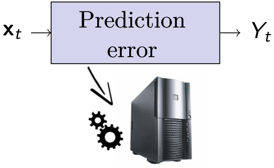


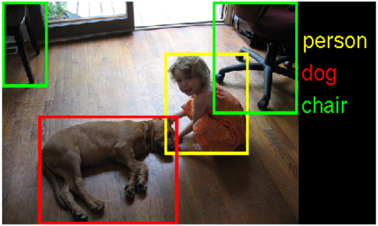




Example: **Deep Neural Network** for object recognition.







Parameters to tune: Number of neurons (Integer-Valued), number of layers (Integer-Valued), learning-rate, activation function (Categorical-Valued), etc.

Optimization Problems: Common Features

- Very expensive evaluations.

Optimization Problems: Common Features

- Very expensive evaluations.

- The objective is a black-box.

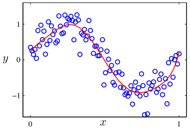
Optimization Problems: Common Features

- Very expensive evaluations.

- The objective is a black-box.

- The evaluation can be noisy.

$$y = f(\mathbf{x}) + \epsilon$$



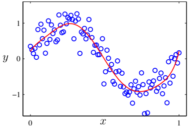
Optimization Problems: Common Features

- Very expensive evaluations.

- The objective is a black-box.

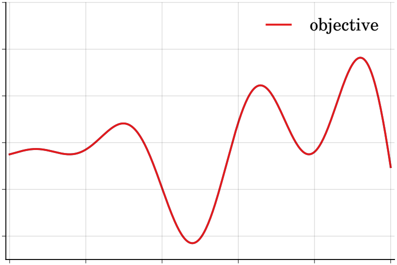
- The evaluation can be noisy.

$$y = f(\mathbf{x}) + \epsilon$$



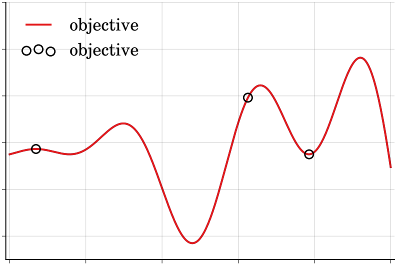
Bayesian optimization methods can be used to solve these problems!

Bayesian Optimization in Practice



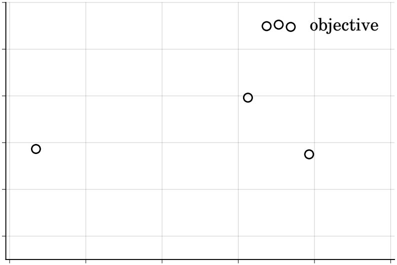
1. Get initial sample.

Bayesian Optimization in Practice



1. Get initial sample.

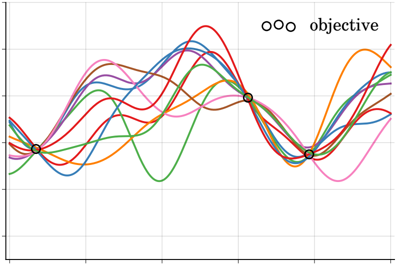
Bayesian Optimization in Practice



1. Get initial sample.
2. Fit a model to the data:

$$p(y|\mathbf{x}, \mathcal{D}_n) \,.$$

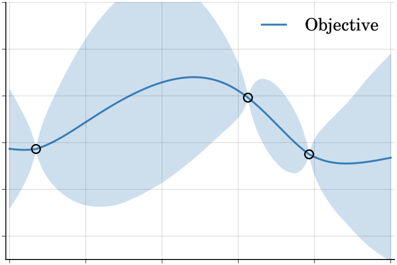
Bayesian Optimization in Practice



1. Get initial sample.
2. Fit a model to the data:

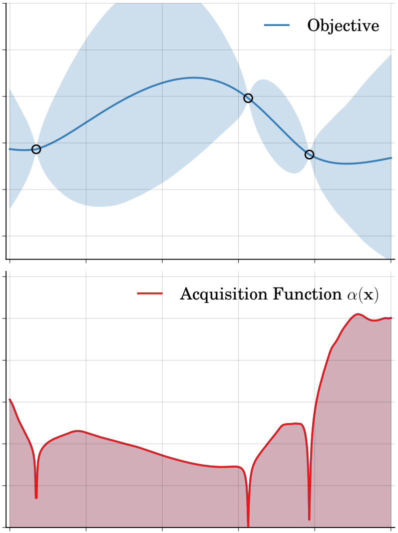
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$

Bayesian Optimization in Practice



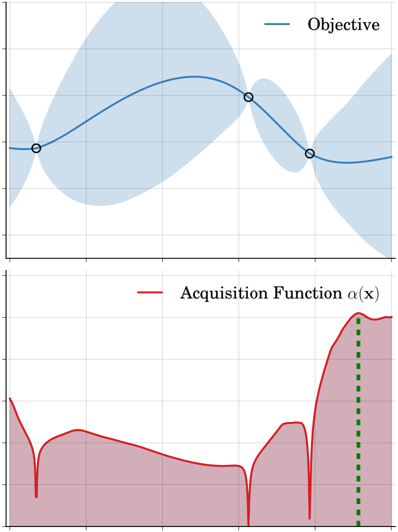
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$

Bayesian Optimization in Practice



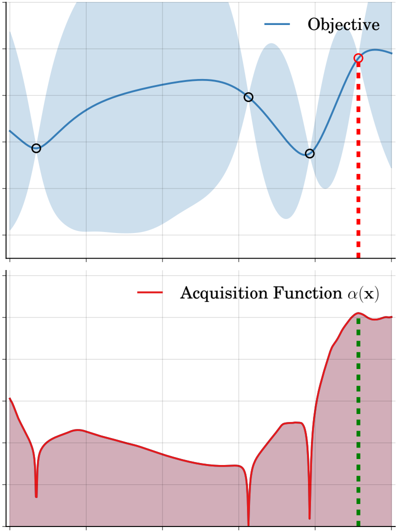
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.

Bayesian Optimization in Practice



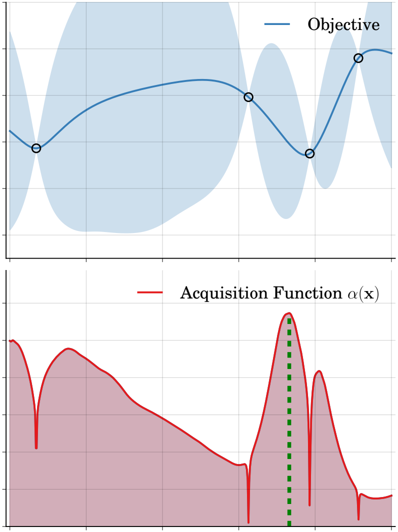
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.

Bayesian Optimization in Practice



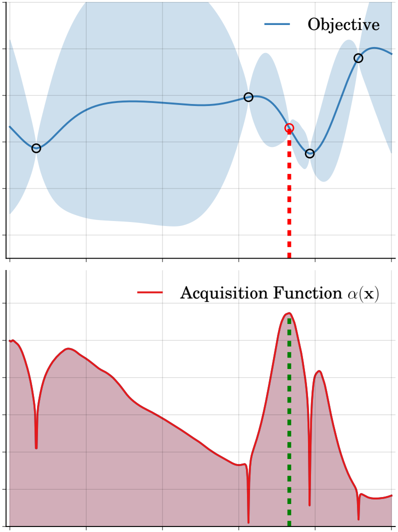
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



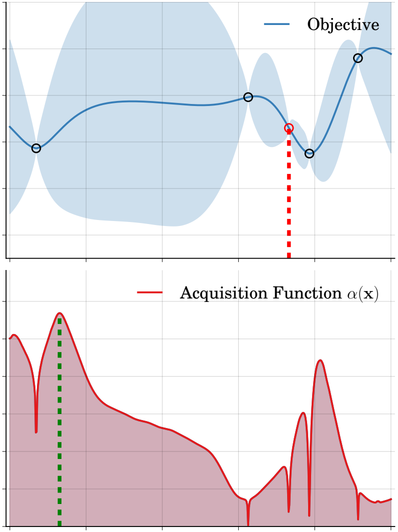
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



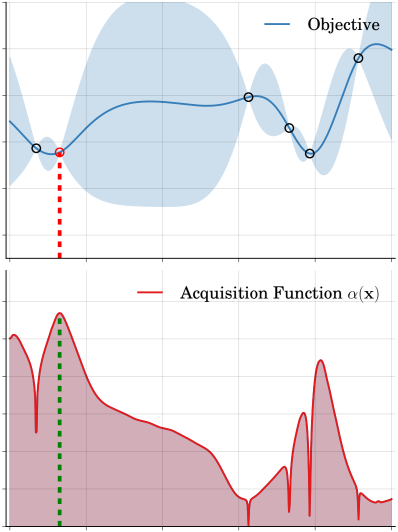
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



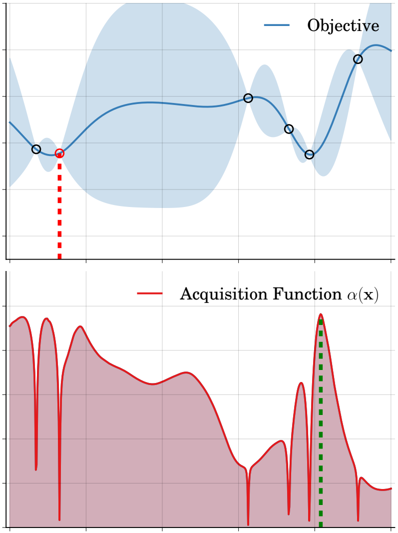
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



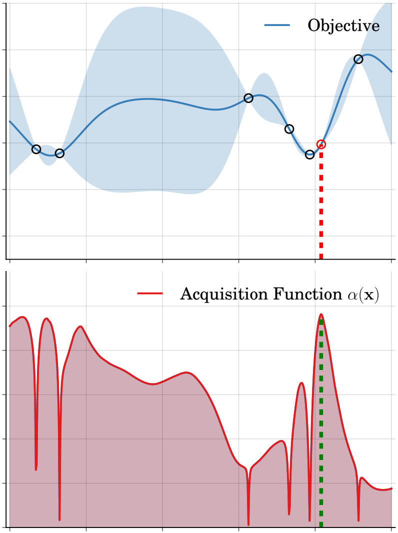
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



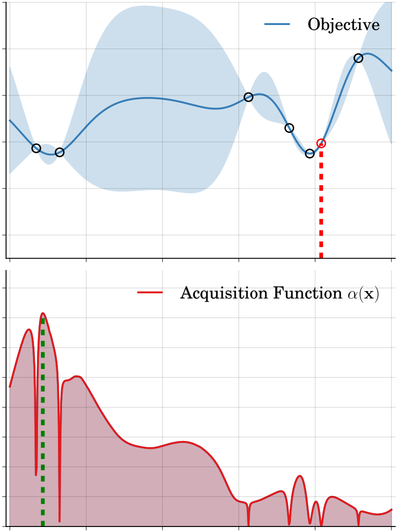
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



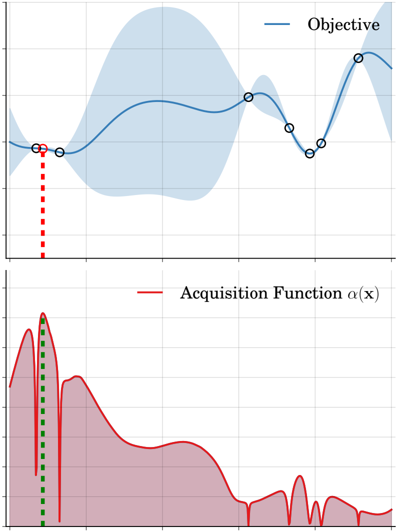
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



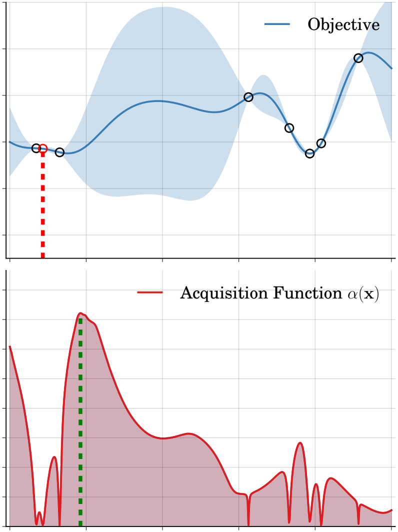
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x}, \mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



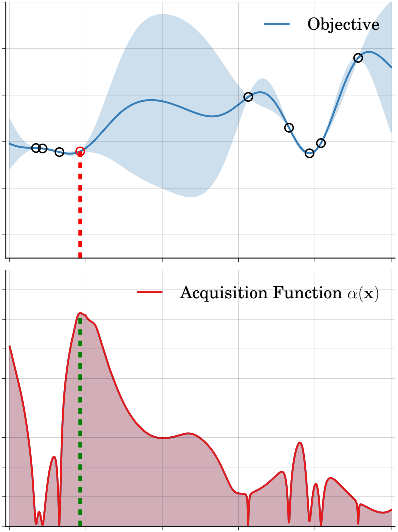
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



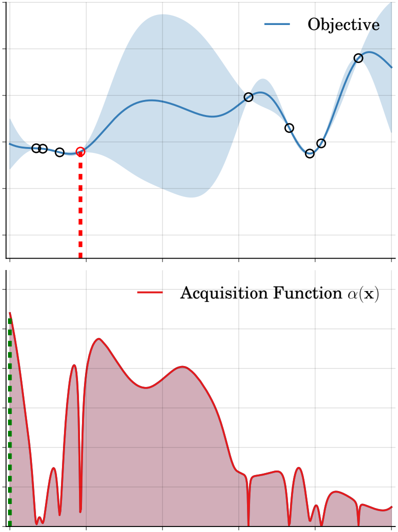
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



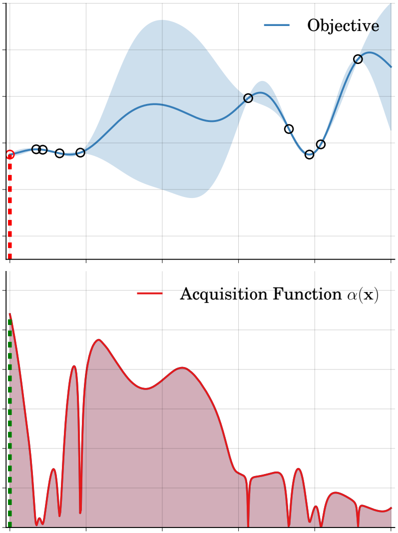
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x}, \mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



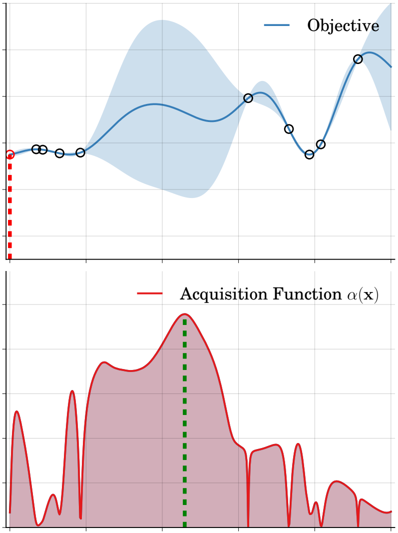
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



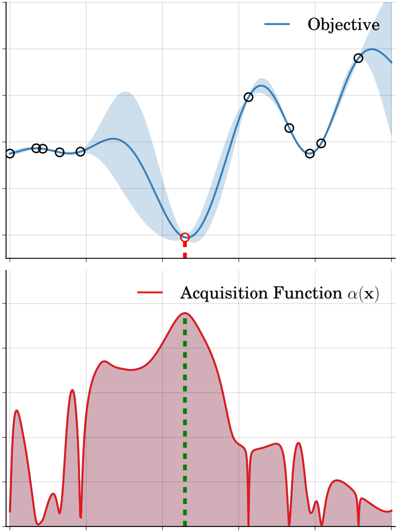
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice



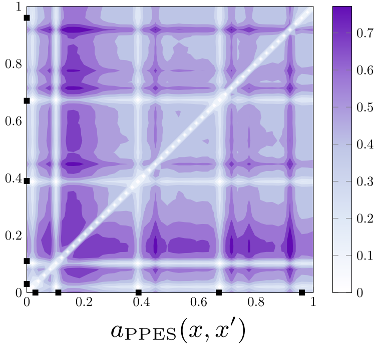
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

Bayesian Optimization in Practice

1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x}, \mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

What if, instead of being built for every point, **the acquisition function is built from a combination of Q points?** Parallel Setup.

Bayesian Optimization in Practice

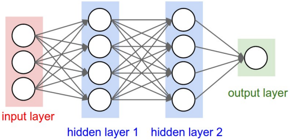


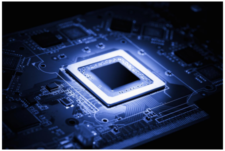
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathbf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathbf{x}) = \mathbb{E}_{p(y|\mathbf{x},\mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathbf{x})$.
5. Collect data and update model.
6. Repeat!

What if, instead of being built for every point, **the acquisition function is built from a combination of Q points?** Parallel Setup.

Several Objectives and Constraints

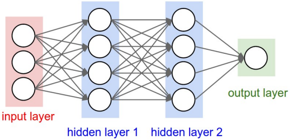
Optimal design of **hardware accelerator** for neural network predictions.

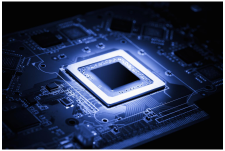




Several Objectives and Constraints

Optimal design of **hardware accelerator** for neural network predictions.



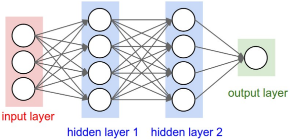


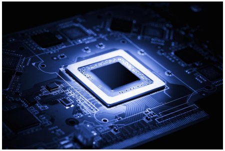
Goals:

- Minimize **prediction error**.
- Minimize **prediction time**.

Several Objectives and Constraints

Optimal design of **hardware accelerator** for neural network predictions.





Goals:

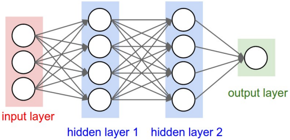
- Minimize **prediction error**.
- Minimize **prediction time**.

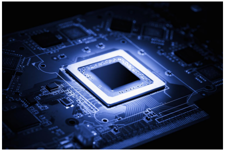
Constrained to:

- **Chip area** below a value.
- **Power consumption** below a level.

Several Objectives and Constraints

Optimal design of **hardware accelerator** for neural network predictions.



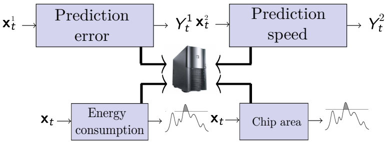


Goals:

- Minimize **prediction error**.
- Minimize **prediction time**.

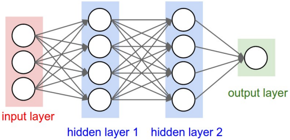
Constrained to:

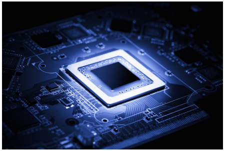
- **Chip area** below a value.
- **Power consumption** below a level.



Several Objectives and Constraints

Optimal design of **hardware accelerator** for neural network predictions.



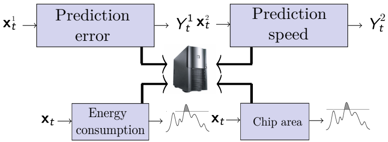


Goals:

- Minimize **prediction error**.
- Minimize **prediction time**.

Constrained to:

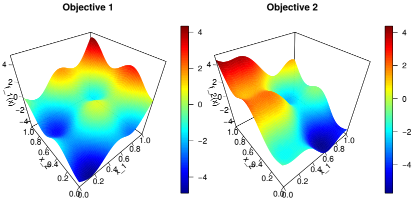
- **Chip area** below a value.
- **Power consumption** below a level.



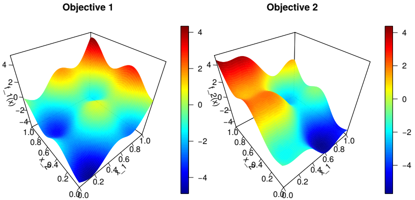
Challenges:

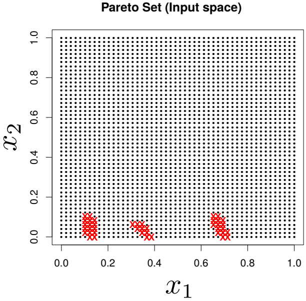
- **Complicated** constraints.
- **Conflictive** objectives.

Constrained Multi-Objective Optimization

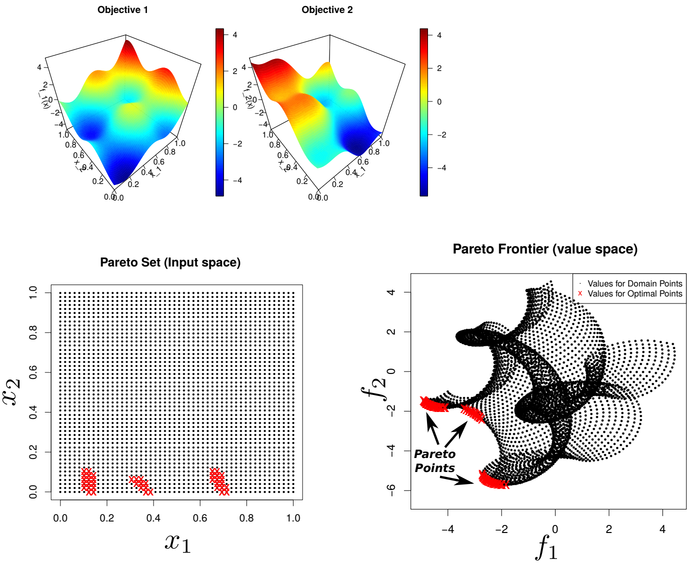


Constrained Multi-Objective Optimization

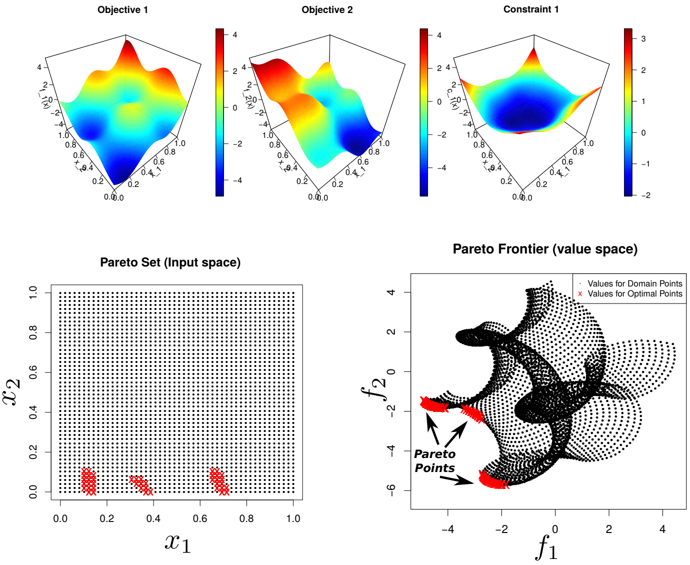




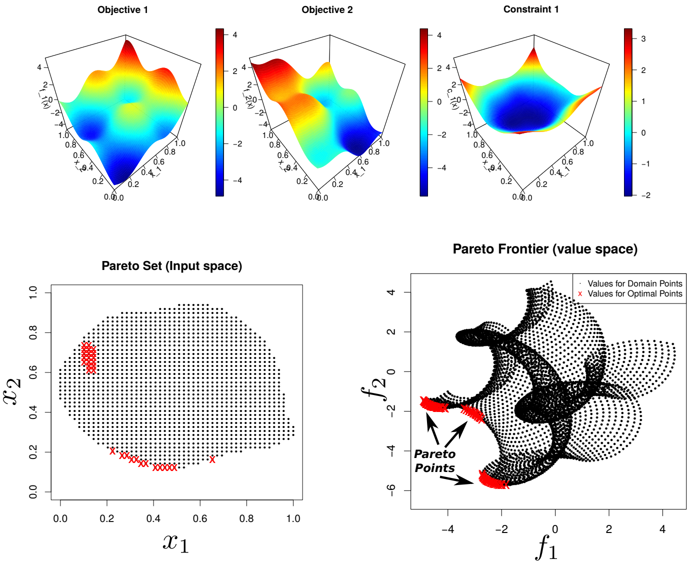
Constrained Multi-Objective Optimization



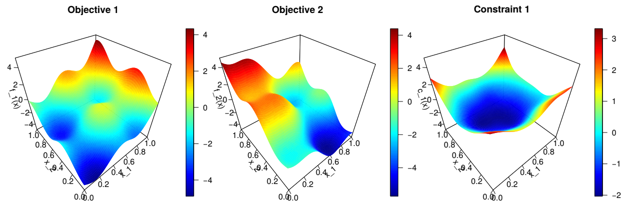
Constrained Multi-Objective Optimization

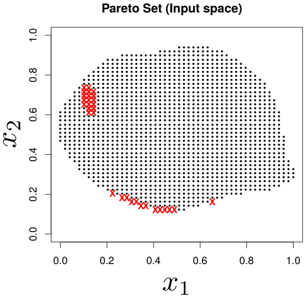


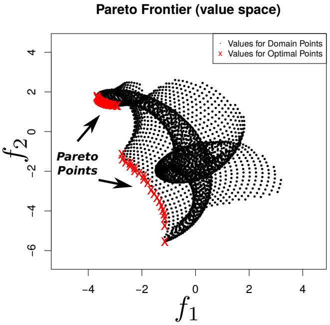
Constrained Multi-Objective Optimization



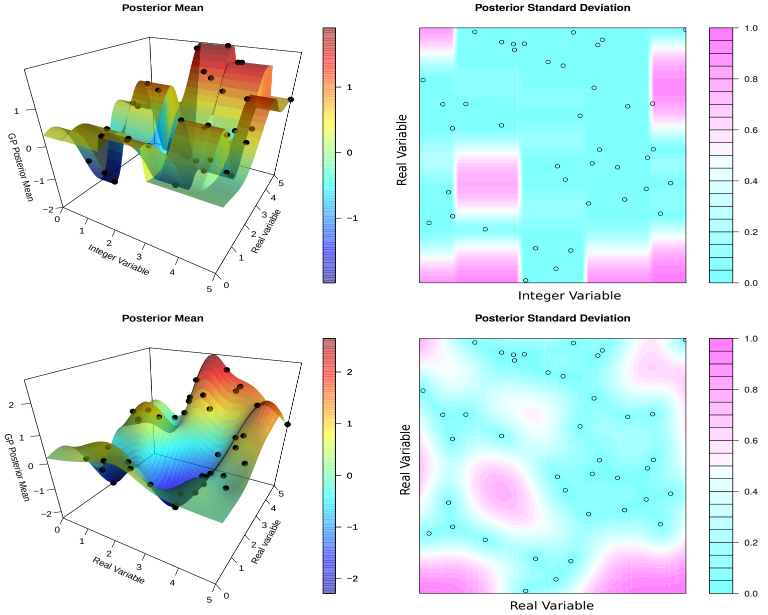
Constrained Multi-Objective Optimization







Visualizing a GP under the *proposed approach*



Applications of Bayesian Optimization

- **Bayesian Networks are Probabilistic Graphical Models** that represent probability distributions. **Gaussian Bayesian Networks represent Multivariate Gaussians**.

Applications of Bayesian Optimization

- **Bayesian Networks are Probabilistic Graphical Models** that represent probability distributions. **Gaussian Bayesian Networks represent Multivariate Gaussians**.
- There exist an exponential space of BN w.r.t nodes that can represent data, I have used BO to reconstruct BN from data searching in that space.

Applications of Bayesian Optimization

- **Bayesian Networks are Probabilistic Graphical Models** that represent probability distributions. **Gaussian Bayesian Networks represent Multivariate Gaussians**.
- There exist an exponential space of BN w.r.t nodes that can represent data, I have used BO to reconstruct BN from data searching in that space.
- I have used BO in an optimization of a Genetic Algorithm to optimize the wave energy retrieved in a real experiment in USA.

Applications of Bayesian Optimization

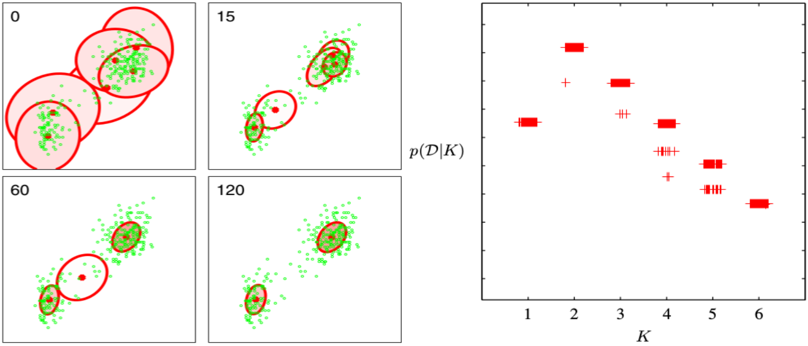
- **Bayesian Networks are Probabilistic Graphical Models** that represent probability distributions. **Gaussian Bayesian Networks represent Multivariate Gaussians**.
- There exist an exponential space of BN w.r.t nodes that can represent data, I have used BO to reconstruct BN from data searching in that space.
- I have used BO in an optimization of a Genetic Algorithm to optimize the wave energy retrieved in a real experiment in USA.
- I have also applied it in an undergraduate thesis in cooking recipes with amazing results!

Ideas for DarkMachines unsupervised learning project

- An initial approach would be to use a **Variational Mixture of Gaussians** to see how many clusters of different objects appear by using this technique.

Ideas for DarkMachines unsupervised learning project

- An initial approach would be to use a **Variational Mixture of Gaussians** to see how many clusters of different objects appear by using this technique.



- **Bayesian inference** automatically makes the **trade-off between model complexity and fitting the data**. **No overfitting** by considering a large K and **maximizing the lower bound**!

Ideas for DarkMachines unsupervised learning project

- This technique could be refined using **Bayesian Optimization** having in count not only one but several objectives and constraints.

Ideas for DarkMachines unsupervised learning project

- This technique could be refined using **Bayesian Optimization** having in count not only one but several objectives and constraints.
- **Gaussian Processes** have also been applied for unsupervised learning with amazing results with the **Gaussian Process Latent Variable Model** (and related approaches) that I think that would be a more advanced approach than the **Variational Mixture of Gaussians**.

References

- Proposed initial Approach: **Variational Mixture of Gaussians**. Nasrabadi, Nasser M. "Pattern recognition and machine learning." Journal of electronic imaging 16.4 (2007): 049901. Section 10.2. Recommended: Read chapter 10.

References

- Proposed initial Approach: **Variational Mixture of Gaussians**. Nasrabadi, Nasser M. "Pattern recognition and machine learning." Journal of electronic imaging 16.4 (2007): 049901. Section 10.2. Recommended: Read chapter 10.
- Proposed Approach: **Gaussian Process Latent Variable Model**. Lawrence, Neil. "Probabilistic non-linear principal component analysis with Gaussian process latent variable models." Journal of machine learning research 6.Nov (2005): 1783-1816.

References

- Proposed initial Approach: **Variational Mixture of Gaussians**. Nasrabadi, Nasser M. "Pattern recognition and machine learning." Journal of electronic imaging 16.4 (2007): 049901. Section 10.2. Recommended: Read chapter 10.
- Proposed Approach: **Gaussian Process Latent Variable Model**. Lawrence, Neil. "Probabilistic non-linear principal component analysis with Gaussian process latent variable models." Journal of machine learning research 6.Nov (2005): 1783-1816.
- Enhanced Proposed Approach, **Bayesian Gaussian Process Latent Variable Model**. Titsias, Michalis, and Neil D. Lawrence. "Bayesian Gaussian process latent variable model." Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 2010.

References

- Utility to consider to decide between approaches or optimize hyperparameters or other spaces, **Bayesian Optimization**. (Feel free to ask me how to apply it, it is my area of expertise). Brochu, Eric, Vlad M. Cora, and Nando De Freitas. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning." arXiv preprint arXiv:1012.2599 (2010).

References

- Utility to consider to decide between approaches or optimize hyperparameters or other spaces, **Bayesian Optimization**. (Feel free to ask me how to apply it, it is my area of expertise). Brochu, Eric, Vlad M. Cora, and Nando De Freitas. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning." arXiv preprint arXiv:1012.2599 (2010).
- Bayesian Deep Learning proposed approach: **Probabilistic backpropagation for scalable learning of bayesian neural networks**. Herndez-Lobato, Jos Miguel, and Ryan Adams. "Probabilistic backpropagation for scalable learning of bayesian neural networks." International Conference on Machine Learning. 2015.

References

- Utility to consider to decide between approaches or optimize hyperparameters or other spaces, **Bayesian Optimization**. (Feel free to ask me how to apply it, it is my area of expertise). Brochu, Eric, Vlad M. Cora, and Nando De Freitas. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning." arXiv preprint arXiv:1012.2599 (2010).
- Bayesian Deep Learning proposed approach: **Probabilistic backpropagation for scalable learning of bayesian neural networks**. Herndndez-Lobato, Jos Miguel, and Ryan Adams. "Probabilistic backpropagation for scalable learning of bayesian neural networks." International Conference on Machine Learning. 2015.
- Background, **Gaussian Processes**. Rasmussen, Carl Edward. "Gaussian processes in machine learning." Advanced lectures on machine learning. Springer, Berlin, Heidelberg, 2004. 63-71.