Chapter 9. Bayesian Decision Theory

Definition

A Bayesian statistician is an individual who makes decisions based on the probability distribution of those things we don't know conditional on what we know, i.e.

$$p(\theta|y, K).$$

Bayesian decision theory

Suppose we have an unknown quantity θ which we believe follows a probability distribution $p(\theta)$ and a decision (or action) δ. For each decision, we have a loss function $L(\theta, \delta)$ that describes how much we lose if θ is the truth. The expected loss is taken with respect to $\theta \sim p(\theta)$, i.e.

$$E_\theta[L(\theta, \delta)] = \int L(\theta, \delta) p(\theta) d\theta = f(\delta).$$

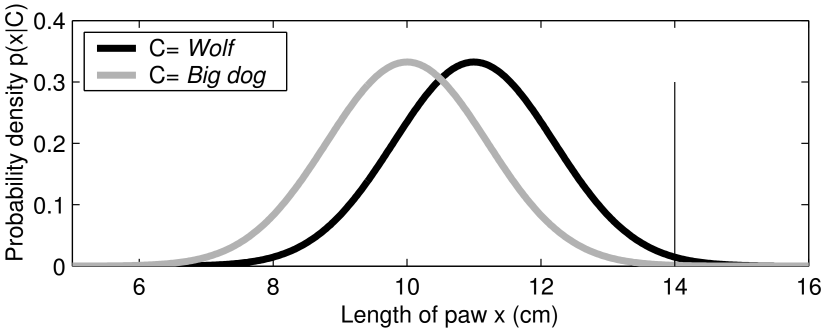
The optimal Bayesian decision is to choose δ that minimizes the expected loss, i.e.

$$\delta_{opt} = \text{argmin}_\delta E[L(\theta, \delta)] = \text{argmin}_\delta f(\delta).$$

Economists typically maximize expected utility where utility is the negative of loss, i.e. $U(\theta, \delta) = -L(\theta, \delta)$. If we have data, just replace the prior $p(\theta)$ with the posterior $p(\theta|y)$.

Example of decision making: 2 choices

- Helen is going to pick mushrooms in a forest, while she notices a paw print which could made by a dog or a wolf
- Helen measures that the length of the paw print is 14 cm and goes home to Google how big paws dogs and wolves have, and tries then to infer which animal has made the paw print



length has been marked with a line

- Likelihood of wolf is 0.92 (alternative being dog)

Example of decision making

- Helen assumes also that in her living area there are about one hundred times more free running dogs than wolves, that is *a priori* probability for wolf, before observation is 1%.
- Likelihood and posterior

Animal	Likelihood	Posterior probability
Wolf	0.92	0.10
Dog	0.08	0.90

- Posterior probability of wolf is 10%

Example of decision making

- Helen has to make decision whether to go pick mushrooms
- If she doesn't go to pick mushrooms utility is zero
- Helen assigns positive utility 1 for getting fresh mushrooms
- Helen assigns negative utility -1000 for a event that she goes to the forest and wolf attacks (for some reason Helen assumes that wolf will always attack)

	Animal	
Decision δ	Wolf	Dog
Stay home	0	0
Go to the forest	-1000	1

Utility matrix $U(x)$

Action δ	Expected utility $E[U(\theta, \delta)]$
Stay home	0
Go to the forest	-100+0.9 = -99.1

Utilities for different actions

Example of decision making

- Maximum likelihood decision (estimation) would be to assume that there is a wolf
- Maximum posterior decision (estimation) would be to assume that there is a dog
- Maximum utility decision is to stay home, even if it is more likely that the animal is dog
-

Action δ	Expected utility $E[U(\theta, \delta)]$
Stay home	0
Go to the forest	-100+0.9 = -99.1

Parameter estimation (replace δ by $\hat{\theta}$)

Definition

For a given loss function $L(\theta, \hat{\theta})$ where $\hat{\theta}$ is an estimator for θ, the Bayes estimator is the function $\hat{\theta}$ that minimizes the expected loss, i.e.

$$\hat{\theta} = \text{argmin}_{\hat{\theta}}\ E_{\theta|y}\left[L\left(\theta, \hat{\theta}\right) \middle| y \right].$$

Recall that

- $\hat{\theta} = E[\theta|y]$ minimizes $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$
- $0.5 = \int_{-\infty}^{\hat{\theta}} p(\theta|y)d\theta$ minimizes $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$
- $\hat{\theta} = \text{argmax}_\theta p(\theta|y)$ is found as the minimizer of the sequence of loss functions $L(\theta, \hat{\theta}) = -(|\theta - \hat{\theta}| < \epsilon)$ as $\epsilon \to 0$

How many quarters in the jar?

Suppose a jar is filled up to a pre-specified line. Let θ be the number of quarters in the jar. Provide a probability distribution for your uncertainty in θ. Suppose you choose

$$\theta \sim N(\mu, \sigma^2)$$

Since $\theta \in \mathbb{N}^+$, we can provide a formal prior by letting

$$P(\theta = q) \propto N(q; \mu, \sigma^2) I(0 < q \leq U)$$

for some upper bound U.

Guessing how many quarters are in the jar.

Now you are asked to guess how many quarters are in the jar.
What should you guess?
If your guess is right, you will win all the quarters in the jar, then what should you guess?

Let q be the guess that the number of quarters is q, then our utility is

$$U(\theta, q) = qI(\theta = q)$$

and our expected utility is

$$E_\theta[U(\theta, q)] = qP(\theta = q) \propto qN(q; \mu, \sigma^2)I(0 \leq q \leq U).$$

Deriving the optimal decision

Here are three approaches for deriving the optimal decision:

$$\arg\max_q f(q), \quad f(q) = qN(q; \mu, \sigma^2)I(0 \leq q \leq U)$$

1. Evaluate $f(q)$ for $q \in \{1, 2, \ldots, U\}$ and find which one is the maximum.
2. Treat q as continuous and use a numerical optimization routine.
3. Take the derivative of $f(q)$, set it equal to zero, and solve for q.

In all cases, you are better off taking the $\log f(q)$ which is monotonic and therefore will still provide the same maximum as $f(q)$.

Derivation

The function to maximize is

$$\log f(q) = \log(q) - (q-\mu)^2/2\sigma^2.$$

The derivative is

$$\frac{d}{dq}\log f(q) = \frac{1}{q} - (q-\mu)/\sigma^2.$$

Setting this equal to zero and multiplying by $-q\sigma^2$ results in

$$q^2 - \mu q - \sigma^2 = 0.$$

This is a quadratic with roots at

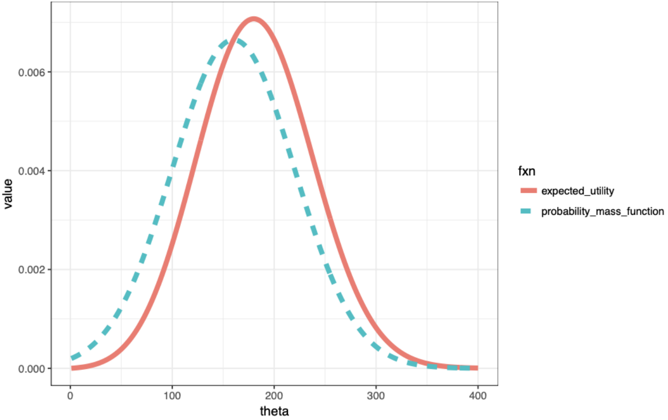
$$\frac{\mu \pm \sqrt{\mu^2 + 4\sigma^2}}{2}.$$

Since q must be positive, the answer is

$$\frac{\mu + \sqrt{\mu^2 + 4\sigma^2}}{2}.$$

Visualizing the utility

Assume $\mu = 160 \quad \sigma = 60 \quad U = 400$



Sequential decisions

Consider a sequence of posteriors distributions $p(\theta_t|y_{1:t})$ that describe your uncertainty about the current state of the world θ_t given the data up to the current time $y_{1:t} = (y_1, \ldots, y_t)$. You also have a loss function for the current time $L(\theta_t, \delta_t)$. Now suppose you are allowed to make a decision δ_{t+1} at each time t and this decision can affect the future states of the world θ_s for $s > t$.

At each time point, we have an optimal Bayes decision, i.e.

$$\text{argmin}_{\delta_{t+1}} \sum_{s=t+1}^{\infty} E_{\theta_s, \delta_s | y_{1:t}} \left[L\left(\theta_s, \delta_s\right) \middle| \, y_{1:t} \right].$$

But because your decision can affect future states which, in turn, can affect future decisions, your current decision needs to integrate over future decisions.

Multi-stage decision making (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - expected lifetime is 34.8 months if no cancer
  - expected lifetime is 16.7 months if cancer and radiation therapy is used
  - expected lifetime is 20.3 months if cancer and surgery, but the probability of dying in surgery is 35% (for 95 year old)
  - expected lifetime is 5.6 months is cancer and no treatment
- Which treatment to choose?
  - quality adjusted life time
  - 1 month is subtracted for the time spent in treatments
- Quality adjusted life time
  - Radiothreapy: 0.9*16.7 + 0.1*34.8 - 1 = 17.5mo
  - Surgery: 0.35*0 + 0.65*(0.9*20.3 + 0.1*34.8 - 1) = 13.5mo
  - No treatment: 0.9*5.6 + 0.1*34.8 = 8.5mo
- See the book for continuation of the example with additional test for cancer

Challenges in decision making

- Non-linear utility functions
- Hard to quantify utility function: What is the cost of human life?
- Multiple parties having different utilities

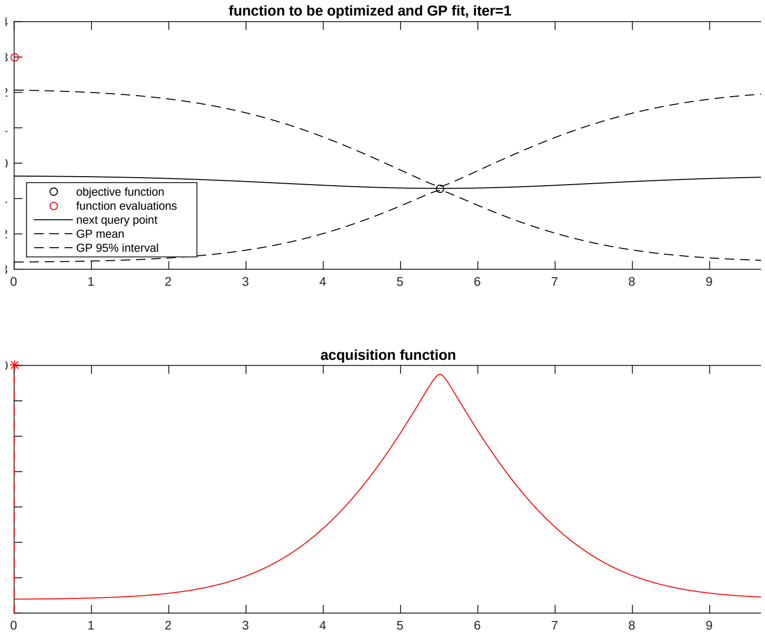
Design of experiment

- Which experiment would give most additional information
  - decide values x_{n+1} for the next experiment
  - which values of x_{n+1} would reduce the posterior uncertainty most
- Example
  - Imagine that in bioassay the posterior uncertainty of LD50 is too large
  - which dose should be used in the next experiment to reduce the variance of LD50 as much as possible ?
    - In this way less experiments need to be made (and less animals need to be killed)

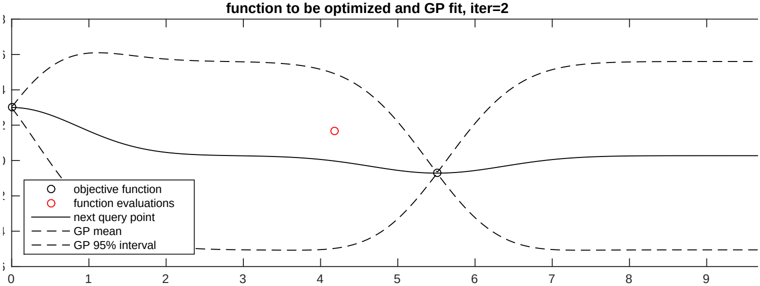
Bayesian optimization

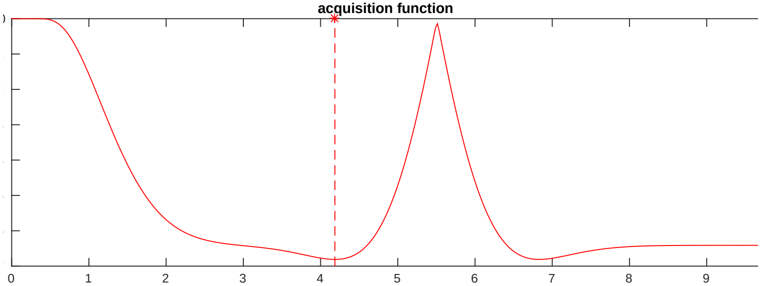
- Design of experiment
- Used to optimize, for example,
  - machine learning / deep learning model structures (hyperparemeters), regularization, and learning algorithm parameters
  - material science
  - engines
  - drug testing

Bayesian optimization

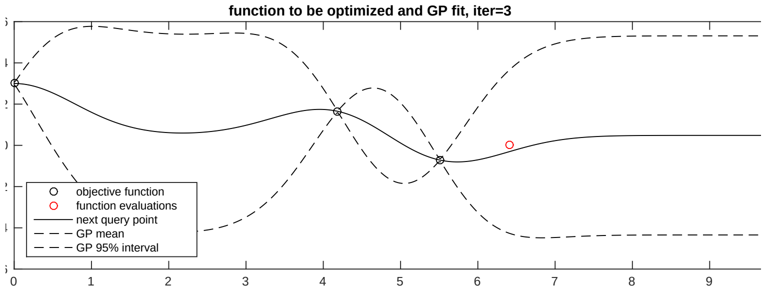


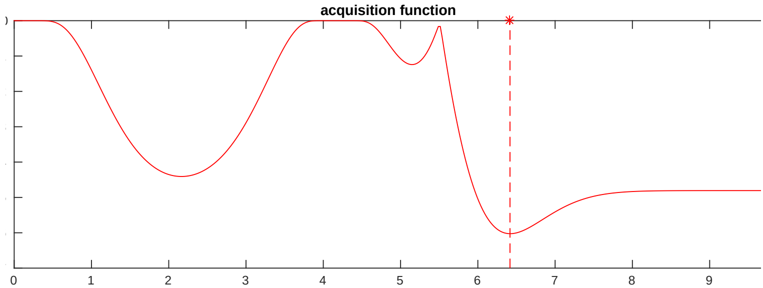
Bayesian optimization



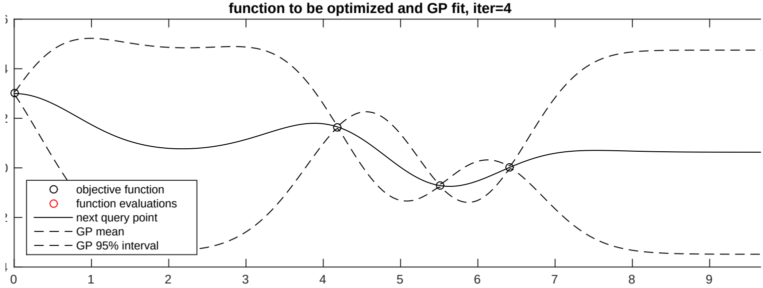


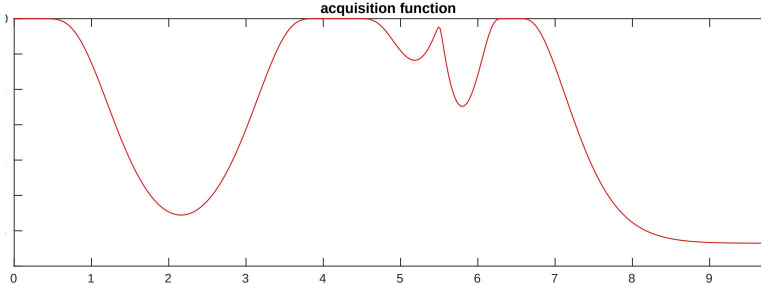
Bayesian optimization



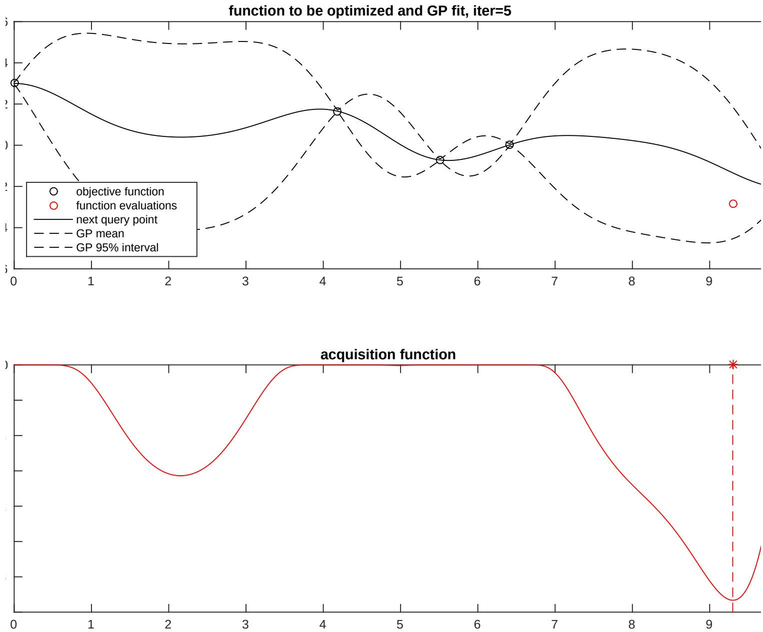


Bayesian optimization

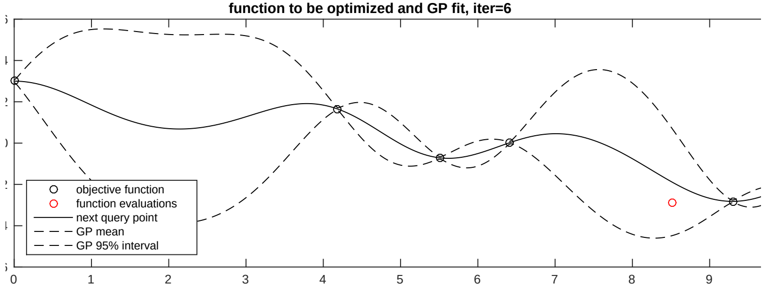


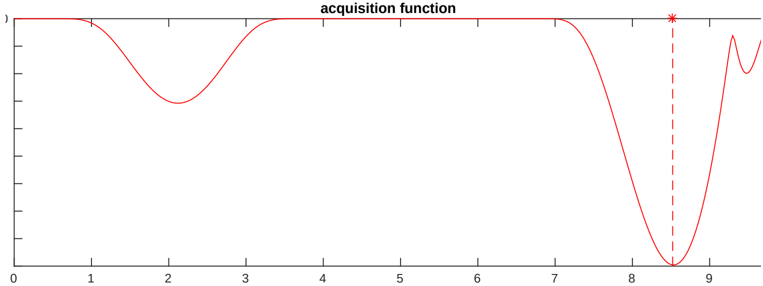


Bayesian optimization

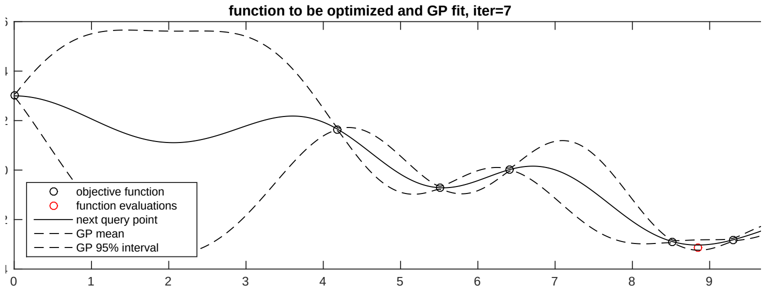


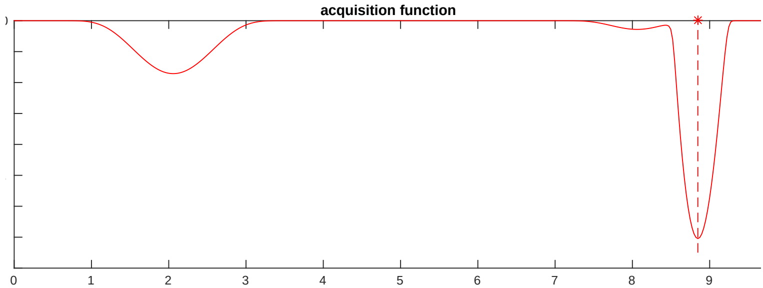
Bayesian optimization



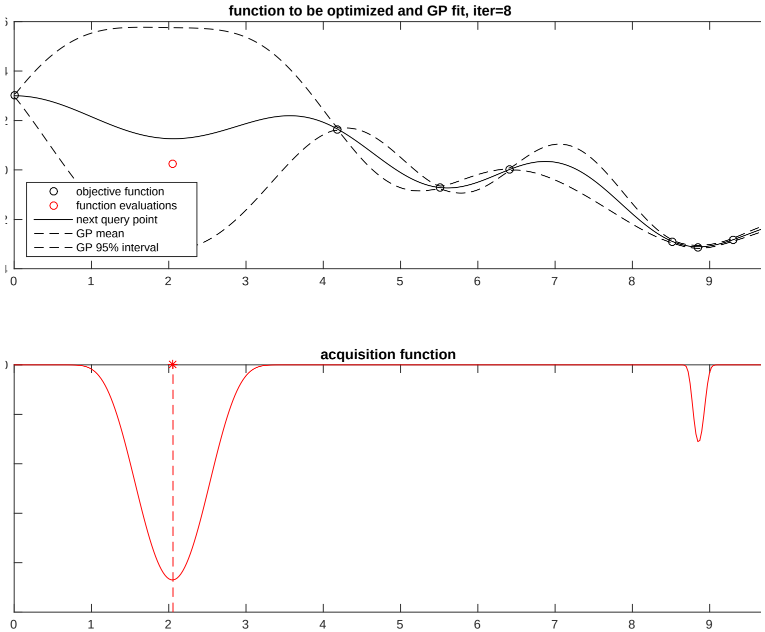


Bayesian optimization

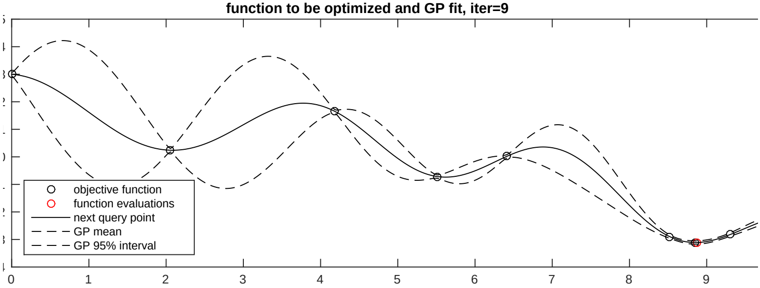


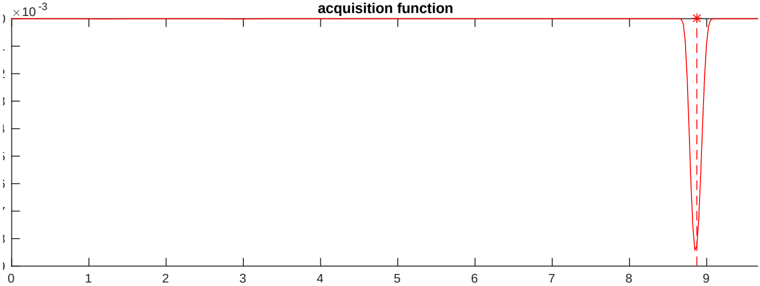


Bayesian optimization

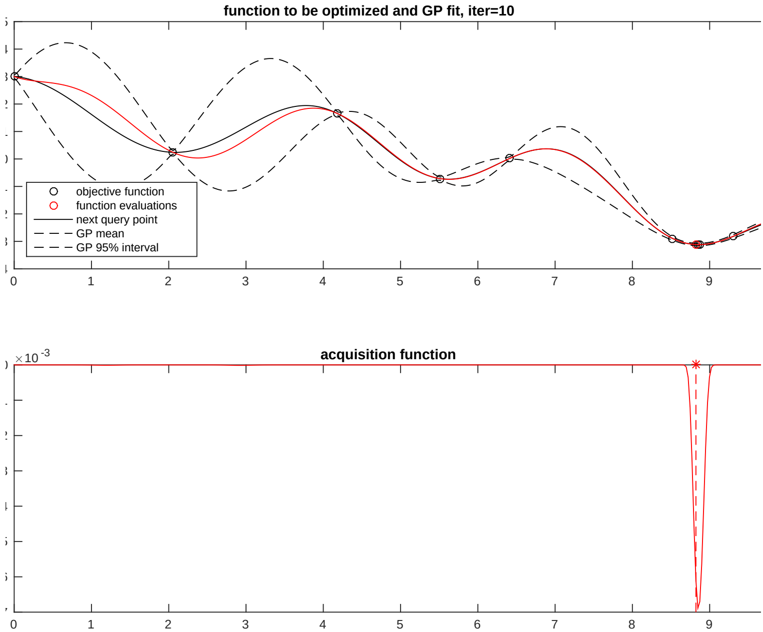


Bayesian optimization





Bayesian optimization



Model selection as decision problem

- Expected utility of using the model in the future