Bayesian inference main idea

Given prior parameter distribution

$$P(\theta)$$

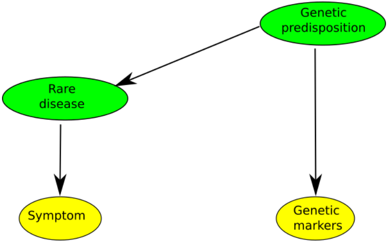
Find its posterior distribution, given training data

$$P(\theta|D)$$

and use it for test data

$$P(D_{\text{test}}|D) = \int P(D_{\text{test}}|\theta, D) \times P(\theta|D) d\theta$$

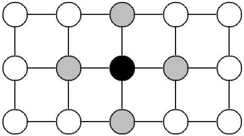
Medical diagnosis problem

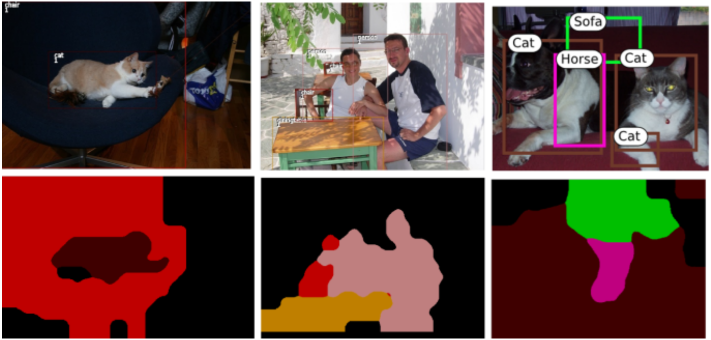


Model = Graph (dependencies) + Conditional Probs

$$P(R\_D = F | S = T, G\_M = F) = \frac{P(S = T | R\_D = F)}{P(S = T | G\_M = F)} \times$$
$$\times \sum_{\alpha:\ G\_P = \alpha} P(R\_D = F | G\_P = \alpha) \times P(G\_P = \alpha | G\_M = F)$$

Markov Random Fields





Flexibility of ML models

There are two ways of achieving flexibility:

1. Large number of parameters compared with dataset (e.g. neural network can have 300 millions); Fit the parameters
2. Non-parametric model. Complexity grows with the amount of trainig data. Fit the data.

Gaussian process model

Definition

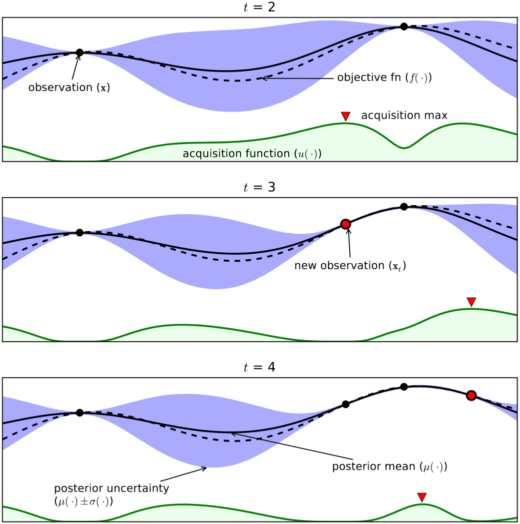
$\{F(x)\}$, $x \in R$, is called Gaussian Process, if for any sequence $x_1 < x_2 < \ldots < x_m$ vector $(F(x_1), F(x_2), \ldots, F(x_m))$ has multivariate normal distribution.

$\{F(x)\}$ is specified by $(m(x); K(x_1, x_2))$ — mean and covariance functions.

Priors:

$$m(x) = m,$$
$$K(x_1, x_2) = \theta_0 \exp\left(-\frac{1}{2}(x_1 - x_2)^2/\theta^2\right).$$

Bayesian optimization



Bayesian optimization

Given training data calculate posterior Gaussian Process with

$$\mu(x, \{x_n, y_n\}, \theta), \quad \sigma(x, \{x_n, y_n\}, \theta).$$

Given posterior distribution, calculate next point:

$$x_{\text{next}} = arg\ max_x a(x)$$

where $a(x)$ is an acquisition function, e.g.

$$a_{\text{PI}}(x; \{x_n, y_n\}, \theta) = \Phi(\gamma(x)), \quad \gamma(x) = \frac{f(x)_{\text{best}} - \mu(x, \{x_n, y_n\}, \theta)}{\sigma(x, \{x_n, y_n\}, \theta)}$$

Bayesian optimization

Branin-Hoo function:

$$f(x_1, x_2) = a(x_2 - bx_1^2 + cx_1 - r)^2 + s(1 - t)cos(x_1) + s,$$

where $x_1 \in [-5; 10]$ and $x_2 \in [0; 15]$.