Machine Learning Working Notes

Abstract

Machine learning is a fast pacing discipline in many working fields, especially it is now regarded as the most impacting subject in artificial intelligence. In this working notes, I summarize some important notes during my study of machine learning. For the completeness, references are included for readers who are reading this article. Note that this working note is only distributed and shared with author's acknowledge and confirmation. It is not intended as a publishable research paper or tutorial.

1. Gaussian Process Regression

Gaussian process is an important nonparametric regression model which looks for an optimal functional in a space of functions, that minimizes a loss function, although the loss function needs not to be explicitly defined. (Rasmussen, 2006)

Give a training dataset $\mathcal{D} = \{x_i\}_{i=1}^n$ with $x_i \in R^d$, *i.i.d* drawn from certain distribution, we are interested at the predictive distribution of unknown target for the test sample x_*, denoted as f_*. Suppose a prior over $\boldsymbol{y}$ given input $\boldsymbol{X}$ is a n-variable Gaussian distribution,

$$\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}, K(\boldsymbol{X}, \boldsymbol{X}))$$

where $K(\boldsymbol{X}, \boldsymbol{X})$ defines a covariance function over $\boldsymbol{X}$. Therefore the posterior of f_* given the training dataset $\mathcal{D}$ is also a Gaussian.

$$f_* | \boldsymbol{y}, \boldsymbol{X}, x_* \sim \mathcal{N}(\mu_*, \Sigma_*^{-1})$$

where the sufficient statistics can be derived using Bayesian theorem,

$$\mu_* = K_{X_*,X} \left[K_{X,X} + \sigma_n^2 I\right]^{-1} \boldsymbol{y} \tag{1}$$

$$\Sigma_*^{-1} = K_{X_*,X_*} - K_{X_*,X} \left[K_{X,X} + \sigma_n^2 I\right]^{-1} K_{X,X_*} \tag{2}$$

where σ_n^2 is the noise level and $K_{\cdot,\cdot}$ represents a shorthand for covariance matrix.

To obtain the Eq.(1)-(2), we can use the following trick. Given two variables $(\boldsymbol{x}, \boldsymbol{y})$ following a Gaussian distribution,

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} A & C \\ C^T & B \end{bmatrix} \right)$$

Then we have the conditional distribution,

$$\boldsymbol{x}|\boldsymbol{y} \sim \mathcal{N}\left(\mu_x + CB^{-1}(\boldsymbol{y} - \mu_y), A - C^T B^{-1} C\right)$$

References

Rasmussen, Carl Edward. Gaussian processes for machine learning. MIT Press, 2006.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.