1 Regression

Suppose we know there exists a function $y : \mathbb{A} \to \mathbb{T}$, such that $t = y(x, \underline{w})$ for $x \in \mathbb{A}$ and $t \in \mathbb{T}$, and $\underline{w}$ is a vector of parameters. If we are given training set $(\underline{x}, \underline{t})$, how do we find y?

Furthermore, if t is noisy, how do we learn y from $(\underline{x}, \underline{t})$? Let's formalize the idea or noisy as follows, each t_i is observed form a random variable $\mathbb{T}_i$, where $\mathbb{T}_i = y(x_i, \underline{w}) + \epsilon$ and $\epsilon \sim N(0, \beta^{-1})$.

Remark Notice $\mathbb{T}_i$ is indeed a random variable, because y is a deterministic function, and ϵ is a random variable.

Now suppose each observations are independent, we have probability

$$\mathbb{P}(\underline{\mathbb{T}} = \underline{t}) = \prod_{i=1}^{n} \mathbb{P}(\mathbb{T}_i = t_i) = \prod_{i=1}^{n} \mathbb{P}\Big(y(x_i, \underline{w}) + \epsilon = t_i\Big)$$

$$= \prod_{i=1}^{n} \mathbb{P}(\epsilon = t_i - y(x_i, \underline{w}))$$

$$= (\frac{\beta}{2\pi})^{\frac{n}{2}} e^{-\frac{\beta}{2}\sum_{i=1}^{n}(t_i - y(x_i, \underline{w}))^2}$$

Now let $\mathbb{L}(\mathbb{X}_1, ..., \mathbb{X}_n) = (\frac{\beta}{2\pi})^{\frac{n}{2}} e^{-\frac{\beta}{2}\sum_{i=1}^{n}(t_i - y(\mathbb{X}_i, \underline{w}))^2}$ be a statistics with parameter $\underline{w}, y$, we can estimate $\underline{w}, y$ by maximum likelihood method. But since y is a function, it is harder to optimize (in Gaussian process we will see how to optimize w.r.t y), we usually will fix y, and optimize w.r.t $\underline{w}$.

1.1 Linear Regression

In linear regression, we transform x to the same dimension of $\underline{w}$ by a feature function $\phi(x) = \begin{bmatrix} \phi_1(x) \\ \vdots \\ \phi_n(x) \end{bmatrix} \in \mathbb{R}^n$, and we fix our y in such linear form

$$y(x) = \underline{w}^T \cdot \phi(x)$$