> 整理自同学的笔记

Linear Regression

Given a data set $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^{d+1}$ and $y_i \in \mathbb{R}$.

Linear Regression by Least Squares

$$y_i = \mathbf{w}^T \mathbf{x}_i,$$

where $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \cdots, x_{id})^T$ and $\mathbf{w} = (w_0, w_1, \cdots, w_d)^T$.

Average Fitting error is

$$\begin{aligned} L &= \frac{1}{n}\sum_i^n (y_i - \mathbf{w}^T \mathbf{x}_i)^2 \\ &= \frac{1}{n}\|\mathbf{y} - \bar{\mathbf{X}}\mathbf{w}\|^2 \end{aligned}$$

where $\mathbf{y} = (y_i, \cdots, y_n)$ and $\bar{\mathbf{X}} \in \mathbb{R}^{n\times(d+1)}$.

$$0 = \frac{\partial L}{\partial \mathbf{w}}\bigg|_{\mathbf{w}=\hat{\mathbf{w}}} = -\frac{2}{n}\bar{\mathbf{X}}^T(\mathbf{y} - \bar{\mathbf{X}}\hat{\mathbf{w}}) \Rightarrow \hat{\mathbf{w}} = (\bar{\mathbf{X}}^T\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}^T\mathbf{y}$$

$$\hat{\mathbf{y}} = \bar{\mathbf{X}}(\bar{\mathbf{X}}^T\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}^T\mathbf{y}$$

Projection matrix $P = \bar{\mathbf{X}}(\bar{\mathbf{X}}^T\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}^T$ projects an arbitrary vector to the column space of $\bar{\mathbf{X}}$.

Linear Regression by Maximum Likelihood

$$y_i = \mathbf{w}^T\mathbf{x}_i + \epsilon_i,$$

假设$\mathbf{w}$与$\mathbf{x}_i$给定。

Assumption 1: $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, $y_i|\mathbf{x}_i, \mathbf{w}, \sigma \sim \mathcal{N}(\mathbf{w}^T\mathbf{x}_i, \sigma^2)$.

Assumption 2: IID. $P((\mathbf{x}_1, y_2), \cdots, (\mathbf{x}_n, y_n)) = \prod_i P((\mathbf{x}_i, y_i))$.

$$\begin{aligned} L &= P(y_1, \cdots, y_n|\mathbf{x}_1, \cdots, \mathbf{x}_n, \mathbf{w}, \sigma) \\ &= \frac{P((\mathbf{x}_1, y_2), \cdots, (\mathbf{x}_n, y_n)|\mathbf{w}, \sigma)}{P(\mathbf{x}_1, \cdots, \mathbf{x}_n|\mathbf{w}, \sigma)} \\ &= \frac{\prod_{i=1}^n P(\mathbf{x}_i, y_i|\mathbf{w}, \sigma)}{\prod_{i=1}^n P(\mathbf{x}_i|\mathbf{w}, \sigma)} \\ &= \prod_{i=1}^n P(y_i|\mathbf{x}, \mathbf{w}, \sigma) \end{aligned}$$

```
$$
  \begin{align}
  \log L &= \sum_i^n \log (\frac{1}{\sqrt{2 \pi \sigma^2}}\exp\{\frac{1}{2\sigma^2}(y_i-{\bf
  w}^T{\bf x}_i)^2\})\\
  &=-\frac{n}{2}\log 2\pi - n \log \sigma + \frac{1}{2\sigma^2}\sum_i^n(y_i-{\bf w}^T{\bf x}_i)^2
  \end{align}


$$
```

$$
\begin{aligned}
\log L &= \sum_i^n \log(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{1}{2\sigma^2}(y_i - \mathbf{w}^T\mathbf{x}_i)^2\}) \\
&= -\frac{n}{2}\log 2\pi - n\log\sigma \frac{1}{2\sigma^2}\sum_i^n (y_i - \mathbf{w}^T\mathbf{x}_i)^2
\end{aligned}
$$

$$
\begin{aligned}
\log L &= \sum_i^n \log(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{1}{2\sigma^2}(y_i - \mathbf{w}^T\mathbf{x}_i)^2\}) \\
&= -\frac{n}{2}\log 2\pi - n\log\sigma + \frac{1}{2\sigma^2}\sum_i^n (y_i - \mathbf{w}^T\mathbf{x}_i)^2
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{\partial \log L}{\partial \mathbf{w}}\bigg|_{\mathbf{w}=\hat{\mathbf{w}}} \\
&= \frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{x}_i y_i - \mathbf{x}_i\mathbf{x}_i^T\hat{\mathbf{w}} \\
&= \frac{1}{\sigma^2}(\bar{\mathbf{X}}^T\mathbf{y} - \bar{\mathbf{X}}^T\bar{\mathbf{X}}\hat{\mathbf{w}})) \\
&\Rightarrow \hat{\mathbf{w}} = (\bar{\mathbf{X}}^T\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}^T\mathbf{y}
\end{aligned}
$$