



Discriminative Models

Linear Model

$$f(x) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n \rightarrow y$$

$$y = w_1 x_1 + w_2 x_2 + \dots + w_n x_n$$

(prior) $\hat{y} = \underline{w}^T \underline{x}$ ← OLS

Linear Regression → Loss → MSE → $\min \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2$

$$\underline{w}^*_{\text{optimal}} = (\underline{x}^T \underline{x})^{-1} \underline{x}^T \underline{y}$$

$$\hat{y} = \underline{w}^T \underline{x}$$

MLE (estimate $\underline{\theta}(\underline{w})$)

$$\hat{\underline{\theta}} = \underset{\underline{\theta}}{\operatorname{argmax}} L(\underline{\theta})$$

$$\hat{\theta}_{\text{MLE}} = \underset{\underline{\theta}}{\operatorname{argmax}} P(D|\underline{\theta})$$

Assumption: IID

$$P(D|\underline{\theta}) = \prod_{i=1}^N P(y_i|x_i, \underline{\theta})$$

$$= \log \prod_{i=1}^N P(y_i|x_i, \underline{\theta})$$

$$\hat{\theta}_{\text{MLE}} = \sum_{i=1}^N \log P(y_i|x_i, \underline{\theta})$$

$$\text{NLL} = -\underset{\text{min}(NLL)}{\operatorname{argmin}} \sum_{i=1}^N \log P(y_i|x_i, \underline{\theta})$$

→ MAP

$$\begin{aligned} &\text{MLE} \rightarrow \text{Bayes Theorem} \\ &P(Y|X) = \frac{P(X|Y)P(Y)}{\text{evidence}} \end{aligned}$$

MLE for Bernoulli

$$(PDF) P(y_i|\underline{\theta}) = \underline{\theta}^{y_i} (1-\underline{\theta})^{1-y_i}$$

$$\text{NLL} = -\log \sum_{i=1}^N \underline{\theta}^{y_i} (1-\underline{\theta})^{1-y_i}$$

$$= -N_1 \log \underline{\theta} + N_0 \log (1-\underline{\theta})$$

$$N = N_1 + N_0$$

$$\frac{d(\text{NLL})}{d\underline{\theta}} = -\frac{N_1}{\underline{\theta}} + \frac{N_0}{1-\underline{\theta}} \xrightarrow{\text{set } \underline{\theta}}$$

$$\hat{\theta}_{\text{MLE}} = \frac{N_1}{N_1 + N_0}$$

MLE for Normal distribution

$$\text{NLL}(\mu, \sigma^2) \xrightarrow{\frac{\partial(\text{NLL})}{\partial \mu} \xrightarrow{\text{set } \sigma^2}}$$

$$\xrightarrow{\frac{\partial(\text{NLL})}{\partial(\sigma^2)} \xrightarrow{\text{set } \mu} \text{solve for } \sigma^2}$$

Logistic Regression → classification

$$\begin{aligned} \underline{y} &= \underline{w}^T \underline{x} \\ \hat{y} &= \sigma(\underline{w}^T \underline{x}) \end{aligned}$$

$$\sigma(a) = \frac{1}{1+e^{-a}}$$

$$\sigma'(a) = \sigma(a) \cdot (1 - \sigma(a))$$

$$\sigma(-a) = 1 - \sigma(a)$$

$$\text{MLE} P(y|x, \underline{\theta}) = \text{Ber}(y | \sigma(\underline{w}^T \underline{x}))$$

$$\text{NLL} = -\frac{1}{N} \sum_{i=1}^N \log [M_n^{y_i} (1-M_n)^{1-y_i}]$$

$$\text{where } M_n = \sigma(\underline{w}^T \underline{x}), \underline{\theta}$$

$$= -\frac{1}{N} (y_i \log M_i + (1-y_i) \log (1-M_i))$$

$$(\text{cross-entropy}) \quad \min(\text{NLL})$$

Loss function → optimization problem

↓ solve

Gradient Descent

