

# Assignment<sub>1</sub>

Soumyadip Santra

February 2021

## 1 Introduction

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \quad (1)$$

- Data set:  $D = \{y_i, x_i\}_{i=1}^m$
- Input(features):  $x_i \in R^n$ ,  $i = 1, \dots, m$
- Outputs:  $y_i \in \mathcal{Y}$ ,  $i = 1, \dots, m$
- Parameters:  $\theta \in R^n$
- Hypothesis:  $h_\theta(x) = \theta^T x$
- Linear model:  $y_i \approx \theta^T x_i$

$$y_i = \theta^T x_i + \epsilon_i \quad (2)$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad (3)$$

$\epsilon_i \leftarrow$  idenpendent,identically distributed random variable

$$p(\epsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\epsilon_i^2}{2\sigma^2}\right) \quad (4)$$

$$p(y_i - \theta^T x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}\right) \quad (5)$$

$$p(y_i | x_i, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}\right) \quad (6)$$

## 2 Maximum Likelihood Estimation(MLE) for $\theta$ :

$$\begin{aligned}
\theta^* &= \arg \max_{\theta} L(\theta \mid D) \\
&= \arg \max_{\theta} p(D \mid \theta) \\
&= \arg \max_{\theta} p(y_1, x_1, y_2, x_2, \dots, y_m, x_m \mid \theta) \\
&= \arg \max_{\theta} p(y_1, x_1 \mid \theta) p(y_2, x_2 \mid \theta) \dots p(y_m, x_m \mid \theta) \\
&= \arg \max_{\theta} \prod_{i=1}^m p(y_i, x_i \mid \theta) \\
&= \arg \max_{\theta} \prod_{i=1}^m p(y_i \mid x_i, \theta) p(x_i \mid \theta) \\
&= \arg \max_{\theta} \prod_{i=1}^m p(y_i \mid x_i, \theta) p(x_i) \\
&= \arg \max_{\theta} \prod_{i=1}^m p(y_i \mid x_i, \theta) \\
&= \arg \max_{\theta} \sum_{i=1}^m \log p(y_i, x_i \mid \theta) \\
&= \arg \max_{\theta} \sum_{i=1}^m \log \left( \frac{1}{\sqrt{2\pi}\sigma} \right) + \log \left[ \exp \left( -\frac{(y_i - \theta^T x_i)^2}{2\sigma^2} \right) \right] \\
&= \arg \max_{\theta} \sum_{i=1}^m -\frac{1}{2\sigma^2} (y_i - \theta^T x_i)^2 \\
&= \arg \min_{\theta} \frac{1}{m} \sum_{i=1}^m (y_i - \theta^T x_i)^2
\end{aligned} \tag{7}$$

- Cost function:  $E(\theta) = \frac{1}{m} \sum_{i=1}^m (y_i - \theta^T x_i)^2$
- Loss function:  $\ell(h_{\theta}(x), y) = (h_{\theta}(x) - y)^2$

## 3 Estimation with MAP for $\theta$ :

$$p(\theta) = \frac{1}{\sqrt{2\pi}r} \exp \left( -\frac{\theta^T \theta}{2r^2} \right) \tag{8}$$

$$\begin{aligned}
\theta^* &= \arg \max_{\theta} p(\theta \mid D) \\
&= \arg \max_{\theta} p(D \mid \theta) p(\theta) \\
&= \arg \max_{\theta} p(y_1, x_1, y_2, x_2, \dots, y_m, x_m \mid \theta) p(\theta) \\
&= \arg \max_{\theta} p(y_1, x_1 \mid \theta) p(y_2, x_2 \mid \theta) \dots p(y_m, x_m \mid \theta) p(\theta) \\
&= \arg \max_{\theta} \prod_{i=1}^m p(y_i, x_i \mid \theta) p(\theta) \\
&= \arg \max_{\theta} \prod_{i=1}^m p(y_i \mid x_i, \theta) p(x_i \mid \theta) p(\theta) \\
&= \arg \max_{\theta} \prod_{i=1}^m p(y_i \mid x_i, \theta) p(x_i) p(\theta) \\
&= \arg \max_{\theta} \prod_{i=1}^m p(y_i \mid x_i, \theta) p(\theta) \\
&= \arg \max_{\theta} \sum_{i=1}^m \log p(y_i, x_i \mid \theta) + \log p(\theta) \\
&= \arg \max_{\theta} \sum_{i=1}^m \log \left( \frac{1}{\sqrt{2\pi}\sigma} \right) + \log \left[ \exp \left( -\frac{(y_i - \theta^T x_i)^2}{2\sigma^2} \right) \right] + \log \left( \frac{1}{\sqrt{2\pi}r} \right) + \log \left[ \exp \left( -\frac{\theta^T \theta}{2r^2} \right) \right] \\
&= \arg \max_{\theta} \sum_{i=1}^m -\frac{1}{2\sigma^2} (y_i - \theta^T x_i)^2 + -\frac{\theta^T \theta}{2r^2} \\
&= \arg \min_{\theta} \frac{1}{m} \sum_{i=1}^m (y_i - \theta^T x_i)^2 + \lambda \|\theta\|_2^2
\end{aligned} \tag{9}$$

- Cost function:  $E(\theta) = \frac{1}{m} \sum_{i=1}^m (y_i - \theta^T x_i)^2 + \lambda \|\theta\|_2^2$
- Loss function:  $\ell(h_{\theta}(x), y) = (h_{\theta}(x) - y)^2 + \lambda \|\theta\|_2^2$