Introduction to Big Data Analysis

Homework 1 Reference Answer

1. Volume, Variety, Value, Velocity
2. D
3. B, C
4. We have

$$\begin{aligned} J(\theta_1) &= \frac{1}{2\times 3}\left[(h_\theta(x^{(1)})-y^{(1)})^2+(h_\theta(x^{(2)})-y^{(2)})^2+(h_\theta(x^{(3)})-y^{(3)})^2\right] \\ &= \frac{1}{2\times 3}\left[(\theta_1-1)^2+(2\theta_1-2)^2+(3\theta_1-3)^2\right] \\ &= \frac{1}{6}[14(\theta_1-1)^2] \\ &= \frac{7}{3}(\theta_1-1)^2. \end{aligned}$$

   It gives that $J(0)=\frac{7}{3}$.

5. A, D, E, F
6. F, T
7.

1NN
Because we have

$$\begin{aligned} \sum_c p_c(x)(1-p_c(z)) &\xrightarrow{\delta\to 0} \sum_c p_c(x)(1-p_c(x)) \\ &\geq \sum_c p_c(x)(1-p_{c^*}(x)) = 1-p_{c^*}(x) = \text{Bayes error} \end{aligned}$$