CS2109S Tutorial 5

AY 25/26 Sem 1 — github/omgeta

A.

1. (a) No regularization: $w_0 = 1.0, w_1 = 0.5$

   (b) Lasso regression with $\lambda = 5$: $w_0 = 0.0, w_1 = 0.5$

   (c) Ridge regression with $\lambda = 5$: $w_0 = 0.25, w_1 = 0.25$

2. $L1$ zeroes w_0 favouring feature selection and sparsity. $L2$ shrinks weights but doesn't reduce any to 0.

B.

1. $X = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \end{pmatrix} \implies X^T X = \begin{pmatrix} 2 & 3 & 6 \\ 3 & 5 & 10 \\ 6 & 10 & 20 \end{pmatrix} \implies det(X^T X) = 0.$

   By Invertible Matrix Theorem, $X^T X$ is singular and therefore cannot have an inverse making it unsuitable for normal equation.

2. $X^T X + \lambda I = \begin{pmatrix} 3 & 3 & 6 \\ 3 & 6 & 10 \\ 6 & 10 & 21 \end{pmatrix} \implies det(X^T X) \neq 0$. By Invertible Matrix Theorem, $X^T X + \lambda I$ has an inverse making it suitable for normal equation.

3. By MATLAB, $h_w(x) = \begin{pmatrix} \frac{13}{33} \\ \frac{5}{11} \\ \frac{10}{11} \end{pmatrix}^\top x = \frac{13}{33} + \frac{5}{11} x_1 + \frac{10}{11} x_2$

C.

1. Points $2, 3, 4$

2. (a) $b = -\begin{pmatrix} 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -0.5$

   (b) Given $h_w(x) = sign(w^\top x + b)$, $h_w(\begin{pmatrix} 0.2 \\ 0.9 \end{pmatrix}) = 1, h_w(\begin{pmatrix} -0.5 \\ 1 \end{pmatrix}) = -1$

D.

1. $\min\limits_i \dfrac{|w^T x^{(i)} + b|}{||w||}$

2. $\max\limits_{w,b} \min\limits_i \dfrac{|w^T x^{(i)} + b|}{||w||}$

3. No, because by minimizing absolute distance, we lose the sign of the vector relative to the hyperplane.

4. Choose w, b s.t. $\min\limits_i |w^T x^{(i)} + b| = 1$, then we have

   $$\max_{w,b} \min_i \frac{|w^T x^{(i)} + b|}{||w||} = \max_{w,b} \frac{1}{||w||} = \min_{w,b} ||w|| \iff \min_{w,b} \frac{||w||^2}{2}$$