CSCI567 2017 Homework Assignment 1 Solution

Algorithmic component

1 Linear Regression

Question 1.1 $X^{\mathrm{T}}X$ is not invertible

$$r(X^{\mathrm{T}}X) \geq D + 1,$$

where $r(\boldsymbol{M})$ is the rank of matrix $\boldsymbol{M}$.

Question 1.2 Bias solution

$$\begin{aligned}
b^* &= \arg\min_b \|\boldsymbol{y} - b\mathbf{1}_N - X^{\mathrm{T}}\boldsymbol{w}\|^2 && \text{Residual sum of squares} && (1)\\
\mathbf{1}_N^{\mathrm{T}}(\boldsymbol{y} - b^*\mathbf{1}_N - X^{\mathrm{T}}\boldsymbol{w}) &= 0 && \text{Taking derivatives w.r.t } b && (2)\\
b^* &= \frac{1}{N}(\mathbf{1}_N^{\mathrm{T}}\boldsymbol{y} - \mathbf{1}_N^{\mathrm{T}}X^{\mathrm{T}}\boldsymbol{w}) && \text{solve for } b^* && (3)\\
&= \frac{1}{N}\mathbf{1}_N^{\mathrm{T}}\boldsymbol{y} && \frac{1}{N}\sum_n x_{nd} = 0 \Leftrightarrow \mathbf{1}_N^{\mathrm{T}}X^{\mathrm{T}} = \mathbf{0} && (4)
\end{aligned}$$

If the feature values are zero on average, the bias b^* is the average response of training samples.

2 Logistic Regression

Question 2.1 Bias solution

$$\begin{aligned}
b^* &= \min_b -\sum_n \{y_n \log\sigma(b) + (1-y_n)\log[1-\sigma(b)]\} && \text{cross entropy objective} && (5)\\
\sum_n y_n(1-\sigma(b^*)) - (1-y_n)\sigma(b^*) &= 0 && \text{Taking derivatives w.r.t } b && (6)\\
\sigma(b^*) &= \frac{\sum_n y_n}{N} && \text{solve for } b^* && (7)\\
b^* &= \log\left(\frac{\sum_n y_n}{\sum_n(1-y_n)}\right) && && (8)
\end{aligned}$$

$\sigma(b^*)$ is the optimal logistic regression classifier, also the probability that a test sample is labeled as 1, which is the fraction of label 1 samples in the training data.