COMP 9501

Assignment 2 Shen Tianxiang 3030058776

Problem 1:

1.1 (a) No! Because the path $x_1 \to \underline{\underline{x_3}} \leftarrow x_2$ is valid

(b) No! Because the path $x_1 \to x_3 \to \underline{\underline{x_4}} \leftarrow x_3 \leftarrow x_2$ is valid

(c) Yes! Because both x_3 & x_4 are not regulized so neither (a) or (b) is valid.

(d) Yes! Because the only path $x_4 \leftarrow x_2 \leftarrow \underline{\underline{x_1}} \to x_3 \to x_7$ is invalid

(e) No! Because the path $x_4 \leftarrow x_2 \to x_5$ is valid.

1.2 (1) $Prob(X_A, Y) = Prob(X_A | Y) Pr(Y)$

$Prob(X_A, X_B, Y) = Prob(X_A, X_B | Y) Prob(Y)$

$= Prob(X_A | Y) Prob(X_B | Y) Prob(Y)$

$Prob(X_A, X_B, X_C, Y) = Prob(X_A | Y) Prob(X_B | Y) Prob(X_C | Y) Prob(Y)$

(2) Yes, independent. For a spam email, the probability for each of its feature is independent.

(3) $P(Y_j=1 \mid (X_A, X_B, X_C)_j)$

$$= \frac{P[Y_j=1, (X_A, X_B, X_C)_j]}{P[(X_A, X_B, X_C)_j]}$$

$$= \frac{P[(X_A, X_B, X_C)_j \mid Y_j=1]\, P[Y_j=1]}{P[(X_A, X_B, X_C)_j]}$$

$$= \frac{P[X_{Aj} \mid Y_j=1]\, P[X_{Bj} \mid Y_j=1]\, P[X_{Cj} \mid Y_j=1]\, P[Y_j=1]}{P[X_{Aj}, X_{Bj}, X_{Cj}]}$$

Because $P[Y_j=1 \mid (X_A, X_B, X_C)_j] + P[Y_j=0 \mid (X_A, X_B, X_C)_j] = 1$

So we don't need to calculate the constant.

1.3 (a) $l(\pi \mid D) = \log P(D \mid \pi) = \log \pi^{n_{y=1}} (1-\pi)^{n_{y=0}}$

Let $\frac{\partial l}{\partial \pi} = 0 \Rightarrow \pi_{MLE} = \frac{\text{Cont}(Y_i=1)}{n}$

(b) $\mu_{A.1} = \frac{\sum X_{Ai} Y_i}{\text{Cont}(Y=1)}$ $\qquad \mu_{A.0} = \frac{\sum X_{Ai}(1-Y_i)}{\text{Count}(Y=0)}$

(c) $\sigma_{A.1}^2 = \frac{1}{\text{Count}(Y=1)} (X_{Ai} - \mu_{A.1})^2 \cdot (Y_i)$

$\sigma_{A.0}^2 = \frac{1}{\text{Count}(Y=0)} (X_{Ai} - \mu_{A.0})^2 \cdot (1-Y_i)$

Problem 2

2.1 (a) pois: $\mu = e^{\eta}$

(b) Batch gradient: $\theta^{(t+1)} = \theta^{(t)} + \lambda_t X^T (y-\mu)$

(c) Newton's method: $\theta^{(t+1)} = \theta^{(t)} + H^{-1} X^T (y-\mu)$

(d) SG $\theta^{(t+1)} = \theta^{(t)} + \lambda_t x_i (y_i - \mu_i)$

2.2 (a) $l = \sum_{i=1}^{n} \sum_{k=1}^{k} \log h(y_n) + \sum_{i=1}^{n} \sum_{k=1}^{k} (\theta^T x_i y_i - A(\eta_i))$

(b) $\theta^{(t+1)} = \theta^{(t)} + \lambda_t x^T (y-\mu)$

Problem 3

3.1 Linear Regression

1. ① Normal Equation

RSS for traing data set: 98.7293

RSS for testing data set: 115.732

prediction: 128.7213.

② SGD

RSS for training data set: 99.905

RSS for testing data set: 115.1695

prediction: 125.9957

③ RGD

RSS for training dataset: ~~86.2417~~ 98.729

RSS for testing data set: ~~85.35513~~ 115.7316.

prediction: 128.39882

3.2 Logistic Regression.

RSS for training data set: 86.2417

RSS for testing data set: 85.35513