Assignment 4

Wenhe Li wl1508

Problem 1

- 1
  - a
    - Naive Nayes:

$$P(Y=1|X_1,...,X_p) = \frac{P(Y=1)\prod_{i=1}^{p}P(X_i|Y=1)}{\sum_{j=0}^{1}P(Y=j)\prod_{i=1}^{p}P(X_i|Y=1)}$$

    - Logistic Regression:

$$P(Y=1|X_1,...,X_p) = \frac{1}{1+exp(w_0+\sum_{i=1}^{p}w_iX_i)}$$

  - b
    - Naive Bayes

$$ln\frac{P(Y=1)}{P(Y=0)}+\sum_{i=1}^{p}ln\frac{P(x_i|Y=1)}{P(x_i|Y=0)}$$

    - Logistic Regression:

$$w_0+\sum_{i=1}^{p}w_iX_i$$

  - c
    Naive Bayes: $P(X_i|Y=y)$ and $P(Y=y)$
    Logistic Regression: the weight w
  - d
    Naive Bayes:
    We just count the total number of elements and the number of elements that fit certain condition.
    Logistic Regression:
    We use Maximize Conditional likelihood estimation to train on the whole dataset and get the Parameters `W = <w0 ... wn>`
- 2
  Assume, $P(X_j=1|Y=1)=\theta_j$ and $P(X_i=1|Y=0)=\theta_i$,
  we can rewrite the classification for naive bayes by:

$$ln\frac{P(Y=1)}{P(Y=0)}+\sum_{n=1}^{p}ln\frac{\theta_j^n(1-\theta_j)^{1-n}}{\theta_i^n(1-\theta_i)^{1-n}}\Rightarrow w_0+\sum_{n=1}^{p}n(ln\frac{\theta_j}{\theta_i})+\sum_{i=1}^{p}(1-n)(ln\frac{1-\theta_j}{1-\theta_i})$$
$$\Rightarrow w_0+\sum_{n=1}^{p}w_nX_n-\sum_{n=1}^{p}m_n(X_n-1)$$

Problem 2

- 1
  - a

$$\frac{e^x}{1+e^x}=\frac{1}{\frac{1}{e^x}+1}=\frac{1}{e^{-x}+1}$$

  - b

$$\frac{e^{x_c}}{\sum_{c'=1}^{C}e^{x_{c'}}}\Rightarrow\frac{e^{x_c}*e^{\delta}}{e^{\delta}*\sum_{c'=1}^{C}e^{x_{c'}}}\Rightarrow\frac{e^{x_c+\delta}}{\sum_{c'=1}^{C}e^{x_{c'}+\delta}}$$

  - c

    For $y=C$:

$$P(y=C|x,W)=\frac{e^{w_C^Tx}}{\sum_{c'=1}^{C-1}e^{w_{c'}^Tx}+e^{w_C^Tx}}=\frac{1}{1+\sum_{c'=1}^{C-1}e^{w_{c'}^Tx-w_C^Tx}}=$$
$$\frac{1}{1+\sum_{c'=1}^{C-1}e^{(w_{c'}^T-w_C^T)x}}=\frac{1}{1+\sum_{c'=1}^{C-1}e^{v_{c'}^Tx}}$$

    Similiarly, for $y=c$

$$P(y=c|x,W)=\frac{e^{w_c^Tx}}{\sum_{c'=1}^{C-1}e^{w_{c'}^Tx}+e^{w_C^Tx}}=\frac{e^{(w_c^T-w_C^T)x}}{1+\sum_{c'=1}^{C-1}e^{w_{c'}^Tx-w_C^Tx}}=\frac{e^{v_c^Tx}}{1+\sum_{c'=1}^{C-1}e^{v_{c'}^Tx}}$$

  - d

    We can siginificantly reduce the value for $v_{c'}^Tx$ to avoid an overflow while calculation.
- 2
  - a
    - 1 By using log, we can avoid underflow from happening, if the likelihood is extremly small.
    - 2 Also, in the calculation, we might encounter `exp`, which could lead to a overflow. By using log, can avoid it as well.
  - b

    By applying a log on our function, we does not change the `monotonicity` of the function. Therefore, we can still get the classification surface.
- 3

- a

  We define $\delta(Y^l = c) = 1$ and $\delta(Y^l \neq c) = 0$

$$l(W) = ln\prod_{l=1}^{n} P(Y^l|X_l, W) = \sum_{l=1}^{n} lnP(Y^l|X_l, W) = \sum_{l=1}^{n}\sum_{c=1}^{C} \delta(Y^l = c)(w_c^T x - ln\sum_{c'=1}^{C} e^{w_{c'}^t x})$$

- b

$$g_c(W) = \sum_{l=1}^{n} \delta(Y^l = c)(x - \frac{xe^{w_c^T x}}{\sum_{c'=1}^{C} e^{w_{c'}^T x}})$$

Problem 3

Threshold	Epoch	Batch	Init weight scale	Learning Rate	Training Error	Validation Error
30	100	50	.5	.01	5%	10%
25	200	100	.5	.75	.05%	2.4%
30	200	100	.5	.75	.05%	3.5%
25	200	100	.5	.75	.05%	2.4%
25	200	125	.5	.85	.025%	3.6%
25	300	100	.5	.75	.025%	1.9%
25	300	300	.5	.75	.075%	2.7%
25	300	150	.5	.75	.025%	3.5%
25	300	125	.5	.5	.075%	2.8%
27	300	100	.1	.75	.0%	2.5%
27	300	100	.01	.75	.0%	2.4%
27	300	100	.01	.85	.025%	3.2%
27	300	300	.10	.85	.05%	2.7%
27	300	500	.01	.99	.05%	2.6%
27	300	250	.0005	.95	.05%	3.1%
27	300	150	.0001	.9	.0%	2.3%
27	250	100	.009	.95	.05%	2.3%