COMS 4771 Machine Learning (Spring 2015)
Problem Set #1

Anji Zhao - `az2324@columbia.edu`

Discussants: `mo2454`

February 25, 2015

Problem 1

1. Given $S = \left\{ \boldsymbol{x}_{i,1:d}, k_i \right\}_{i=1}^{n}$

$$\text{likelihood } L = \prod_{i=1}^{n} \prod_{j=1}^{d} \prod_{k \in \mathcal{Y}} p_{j,k}^{x_{ij}} (1 - p_{j,k})^{(1 - x_{ij})}$$

$$\ln(L) = \sum_{i=1}^{n} \sum_{j=1}^{d} \sum_{k \in \mathcal{Y}} (x_{ij} \ln(p_{jk}) + (1 - x_{ij}) \ln(1 - p_{jk}))$$

$$\ln(L) = \sum_{i=1}^{n} \sum_{j=1}^{d} \mathbb{1}\{k_i = k\}(\mathbb{1}\{x_j = 1\} \ln(p_{jk}) + (1 - \mathbb{1}\{x_j$$

$$\frac{\partial \ln(L)}{\partial p_{j,k}} = \sum_{i=1}^{n} \sum_{j=1}^{d} \mathbb{1}\{k_i = k\} \Big(\frac{\mathbb{1}\{x_j = 1\}}{p_{jk}} - \frac{1 - \mathbb{1}\{x_j = 1\}}{1 - p_{jk}} \Big)$$

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{d} \mathbb{1}\{k_i = k\} \Big(\frac{\mathbb{1}\{x_j = 1\}}{\hat{p_{jk}}} - \frac{\mathbb{1}\{x_j = 0\}}{1 - \hat{p_{jk}}} \Big)$$

$$\hat{p_{jk}} \sum_{i=1}^{n} \sum_{j=1}^{d} \mathbb{1}\{k_i = k\} \mathbb{1}\{x_j = 0\} = (1 - \hat{p_{jk}}) \sum_{i=1}^{n} \sum_{j=1}^{d} \mathbb{1}\{k_i = k\} \mathbb{1}\{x_j = 1\}$$

$$\hat{p_{jk}} \Big(\sum_{i=1}^{n} \sum_{j=1}^{d} \mathbb{1}\{k_i = k\}(\mathbb{1}\{x_j = 0\} + \mathbb{1}\{x_j = 1\}) \Big) = \sum_{i=1}^{n} \sum_{j=1}^{d} \mathbb{1}\{k_i = k\} \mathbb{1}\{x_j = 1\}$$

$$\hat{p_{jk}} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{d} \mathbb{1}\{k_i = k\} \mathbb{1}\{x_j = 1\}}{\sum_{i=1}^{n} \sum_{j=1}^{d} \mathbb{1}\{k_i = k\}}$$

2. If there are only two newsgroups, we can map $\{0,1\} \mapsto \{1,2\} \mapsto \{-1,+1\}$.

$$f(x) = 1 \text{ when } \Pr(Y=1 \mid X=x) > \Pr(Y=2 \mid X=x).$$

$$\pi_1 \prod_{j=1}^{d} p_{j1}^{x_j}(1-p_{j1})^{(1-x_j)} > \pi_2 \prod_{j=1}^{d} p_{j2}^{x_j}(1-p_{j2})^{(1-x_j)}$$

$$\frac{\prod_{j=1}^{d} p_{j1}^{x_j}(1-p_{j1})^{(1-x_j)}}{\prod_{j=1}^{d} p_{j2}^{x_j}(1-p_{j2})^{(1-x_j)}} > \frac{\pi_2}{\pi_2}$$

$$\prod_{j=1}^{d}\left(\frac{p_{j1}}{p_{j2}}\right)^{x_j}\left(\frac{1-p_{j1}}{1-p_{j2}}\right)^{(1-x_j)} > \frac{\pi_2}{\pi_2}$$

$$\ln\left(\prod_{j=1}^{d}\left(\frac{p_{j1}}{p_{j2}}\right)^{x_j}\left(\frac{1-p_{j1}}{1-p_{j2}}\right)^{(1-x_j)}\right) > \ln\left(\frac{\pi_2}{\pi_2}\right)$$

$$\sum_{j=1}^{d} x_j \ln\left(\frac{p_{j1}}{p_{j2}}\right) + (1-x_j)\ln\left(\frac{1-p_{j1}}{1-p_{j2}}\right) > \ln\frac{\pi_2}{\pi_2}$$

$$\left(\sum_{j=1}^{d} x_j \ln\left(\frac{p_{j1}}{p_{j2}}\right) + (1-x_j)\ln\left(\frac{1-p_{j1}}{1-p_{j2}}\right)\right) - \left(\ln\frac{\pi_2}{\pi_2}\right) > 0$$

$$\sum_{j=1}^{d} x_j\left(\ln\left(\frac{p_{j1}}{p_{j2}}\right) - \ln\left(\frac{1-p_{j1}}{1-p_{j2}}\right)\right) + \ln\left(\frac{1-p_{j1}}{1-p_{j2}}\right) - \ln\frac{\pi_2}{\pi_2} > 0$$

Let $w = \left(\ln\left(\frac{p_{j1}}{p_{j2}}\right) - \ln\left(\frac{1-p_{j1}}{1-p_{j2}}\right)\right)$ and $\theta > \ln\frac{\pi_2}{\pi_2} - \ln\left(\frac{1-p_{j1}}{1-p_{j2}}\right)$.

Let $2 \mapsto -1$ and $1 \mapsto +1$. Then we have $f(x) = 1$ when $(\langle w, x\rangle - \theta) > 0$,

or $\boxed{f(x) = \operatorname{sign}(\langle w, x\rangle - \theta)}$

3. training error $= 0.22163$

   test error $= 0.3760$

Problem 2

Problem 3

Problem 4

Problem 5