CS 4780/5780 Homework 3

Ashutosh Agarwal (ana63), Chengrun Yang (cy438), Yuxiang Peng (yp344), Yue Sun (ys758), Zhi ting chen (zc346)

I. (a) $\nabla L(w) = 2\sum_{i=1}^{n} (w^T x_i - y_i)\, x_i + \lambda \cdot 2w$

Gradient update: ~~w(k+1) = w~~ $\vec{w}' = \vec{w} - c\nabla L(\vec{w}) = \vec{w} - 2c\left[\sum_{i=1}^{n} (\vec{w}^T \vec{x}_i - \vec{y}_i)\vec{x}_i + \lambda \vec{w}\right]$

(b) $\nabla L(w) = 2\sum_{i=1}^{n} (w^T x_i - y_i)\, x_i + \lambda \cdot \mathrm{sgn}(w)$ in which $\mathrm{sgn}(w) = (\mathrm{sgn}(w_1), \mathrm{sgn}(w_2), \cdots, \mathrm{sgn}(w_n))^T$

$\mathrm{sgn}(c) = \frac{|c|}{c}$. $c \in \mathbb{R}$

Update: $w' = w - c\nabla L(w) = w - 2c\sum_{i=1}^{n} (w^T x_i - y_i)\, x_i - \lambda c \cdot \mathrm{sgn}(w)$

(c) $\nabla L(w) = \sum_{i=1}^{n} \frac{-\exp(-y_i w^T x_i) \cdot y_i x_i}{1 + \exp(-y_i w^T x_i)}$

Update: $w' = w - c\nabla L(w) = w + c\sum_{i=1}^{n} \frac{y_i x_i \cdot \exp(-y_i w^T x_i)}{1 + \exp(-y_i w^T x_i)}$

(d) $\nabla L(w) = c\sum_{\substack{i=1 \\ y_i w^T x_i < 1}}^{n} (-y_i x_i) + 2w$ Update: $w' = \cancel{w} - w + c\sum_{\substack{i=1 \\ y_i w^T x_i < 1}}^{n} y_i x_i$

II. a. By induction, $w' = \cancel{w} - w^0 + c\sum_{\substack{i=1 \\ y_i w^T x_i < 1}}^{n} y_i x_i = c\sum_{i=1}^{n} y_i x_i$, which has the form $w^t = \sum_{i=1}^{n} \alpha_i^t x_i$

in which $\alpha_i^t = c y_i$

If ~~k=t~~ w^k has the given form, then

$$w^{t+1} = -w^t + c\sum_{\substack{i=1 \\ y_i w^{tT} x_i < 1}}^{n} y_i x_i = \sum_{i=1}^{n} \alpha_i^{t+1} x_i$$

in which $\alpha_i^{t+1} = \begin{cases} -\alpha_i^t + c y_i, & y_i w^{tT} x_i < 1 \\ -\alpha_i^t, & y_i w^{tT} x_i \geq 1 \end{cases}$

b. $L(\alpha) = c\sum_{i=1}^{n} \max\left[1 - y_i\left(\sum_{j=1}^{n} \alpha_j x_j\right) \cdot x_i, 0\right] + \sum_{i=1}^{n} \alpha_i x_i^T \cdot \sum_{j=1}^{n} \alpha_j x_j$

$= c\sum_{i=1}^{n} \max\left[1 - y_i \sum_{j=1}^{n} \alpha_j K_{ij}, 0\right] + \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j K_{ij}$

c. $$\frac{\partial L(\alpha)}{\partial \alpha_i} = -C \sum_{\substack{j=1 \\ y_j \sum_{i=1}^{n} \alpha_i K_{ij} \leq 1}}^{n} y_j K_{ij} + 2 \sum_{j=1}^{n} \alpha_j K_{ij}$$

Update: $$\alpha_i' = \alpha_i - s \frac{\partial L(\alpha)}{\partial \alpha_i}.$$

III $L(w) = \sum_{i=1}^{n} P_i (w^T x_i - y_i)^2 + \lambda w^T w$

a) $X = [x_1, x_2 \cdots x_n]^T$

$Y = [y_1, y_2 \cdots y_n]^T$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} \qquad \begin{bmatrix} x_{11} & \cdots & x_{1d} \\ x_{n1} & \cdots & x_{nd} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w^T x_1 - y_1 \\ \vdots \\ w^T x_n - y_n \end{bmatrix} = XW - Y$$

$$[w^T x_1 - y_1 \cdots w^T x_n - y_n] \begin{bmatrix} P_1 & & \\ & \ddots & \\ & & P_n \end{bmatrix} \begin{bmatrix} w^T x_1 - y_1 \\ \vdots \\ w^T x_n - y_n \end{bmatrix} = \sum_{i=1}^{n} P_n (w^T x_i - y_i)^2 = (XW - Y)^T P (XW - Y)$$

$\therefore \; L(w) = (XW - Y)^T P (XW - Y) + \lambda w^T w$

b) $$\frac{\partial L(w)}{\partial w} = \frac{\partial [(XW - Y)^T P (XW - Y) + \lambda w^T w]}{\partial w}$$

$$= \frac{\partial [(w^T X^T - Y^T) P (XW - Y) + \lambda w^T w]}{\partial w}$$

$$= \frac{\partial [w^T X^T P X w - w^T X^T P Y - Y^T P X w + Y^T P Y + \lambda w^T w]}{\partial w}$$

$$= (X^T P X) w + (X^T P X)^T w - X^T P Y - Y^T P X + 2\lambda I w$$

$$= 2(X^T P X) w - X^T P Y - Y^T P X + 2\lambda w = 0$$

$$\therefore \; w = \frac{1}{2} (X^T P X + \lambda I)^{-1} (X^T P Y + Y^T P X)$$

IV. Regularized Logistic Regression.

b. RLR:

$$L(w) = \min \frac{1}{n} \sum_{i=1}^{n} \log(1 + e^{-h_w(x_i) y_i}) + \lambda \|W\|_2^2$$

$$= \min \frac{1}{n} \sum_{i=1}^{n} \log(1 + e^{-y_i W^T x_i}) + \lambda \|W\|_2^2$$

$$= \min \frac{1}{N} \sum_{i=1}^{N} \log(1 + e^{-y_i W^T x_i}) + \lambda W^T W.$$

MAP:

$$MAP(w) = \max \prod_{i=1}^{N} P(y_i | x_i, w) \, P(w)$$

$$= \max \prod_{i=1}^{N} \frac{1}{1 + e^{-y_i w^T x_i}} \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{w_j^2}{2\sigma^2}\right)$$

$$\log MAP(w) = \max \sum_{i=1}^{N} -\log(1 + e^{-y_i w^T x_i}) - \log(\sqrt{2\pi}^d \sigma^d) - \frac{\sum_{j=1}^{d} w_j^2}{2\sigma^2}$$

$$= \min \sum_{i=1}^{N} \log(1 + e^{-y_i w^T x_i}) + d \log(\sqrt{2\pi}\,\sigma) + \frac{\sum_{j=1}^{d} w_j^2}{2\sigma^2}$$

$$= \min \frac{1}{N} \sum_{i=1}^{N} \log(1 + e^{-y_i w^T x_i}) + \frac{\sum_{j=1}^{d} w_j^2}{2N\sigma^2}$$

$$\therefore \lambda = \frac{1}{2N\sigma^2}$$

C: MAP:

$$\max \prod_{i=1}^{N} \frac{1}{1+\exp(-y_i W^T x_i)} \times C\exp(-\lambda|W|)$$

$$\log MAP = \max \sum_{i=1}^{N} -\log(1+\exp(-y_i W^T x_i)) + \log C + (-\lambda|W|)$$

$$= \min \sum_{i=1}^{N} \frac{1}{N} \log(1+\exp(-y_i W^T x_i)) + \frac{\lambda}{N}|W|$$

L1 Regularization:

$$L(W) = \min \frac{1}{N} \sum_{i=1}^{N} \log(1+\exp(-y_i W^T x_i)) + \gamma|W|$$

$$\therefore \frac{\lambda}{N} = |\gamma|$$

V Linearity of Naive Bayes.

a) By Bayes rule:

$$P(y=1|\vec{x}) = \frac{P(\vec{x}|y=1)\,P(y=1)}{P(\vec{x})}$$

Naive Bayes Assumption:

$$P(\vec{x}|y) = \prod_{\alpha=1}^{d} P(x_\alpha|y),$$ where x_α is a value for feature α

$$P(\vec{x}) = P(\vec{x}|y=1)\,P(y=1) + P(\vec{x}|y=0)\,P(y=0)$$

Then, by combining everything, we have

$$P(y=1|\vec{x}) = \frac{\prod_{\alpha=1}^{d} P(x_\alpha|y=1)\,P(y=1)}{\prod_{\alpha=1}^{d} P([\vec{x}]_\alpha|y=1)\,P(y=1) + \prod_{\alpha=1}^{d} P([\vec{x}]_\alpha|y=0)\,P(y=0)}$$

b) From part a) we divide both the top and bottom by $\prod_{\alpha=1}^{d} P([x]_\alpha|y=1)\,P(y=1)$

$$P(y=1|\vec{x}) = \frac{1}{1 + \underbrace{\frac{\prod_{\alpha=1}^{d} P([\vec{x}]_\alpha|y=0)\,P(y=0)}{\prod_{\alpha=1}^{d} P([\vec{x}]_\alpha|y=1)\,P(y=1)}}_{= a^{-1}}}$$

$$e^{-\ln a} = (e^{\ln a})^{-1} = a^{-1}$$

$$= \frac{1}{1 + \exp^{-\log}\left(\frac{\prod_{\alpha=1}^{d} P([x]_\alpha|y=1)\,P(y=1)}{\prod_{\alpha=1}^{d} P([x]_\alpha|y=0)\,P(y=0)}\right)}$$

c) From last part we have the following:

$$P(y=1|x) = \frac{1}{1+\exp\left(-\log \frac{\prod_{\alpha=1}^{d} P([x]_\alpha|y=1)\,P(y=1)}{\prod_{\alpha=1}^{d} P([x]_\alpha|y=0)\,P(y=0)}\right)}$$

$$= \frac{1}{1+\exp\left(-\log\frac{P(y=1)}{P(y=0)} - \underbrace{\log \frac{\prod_{\alpha=1}^{d} P([x]_\alpha|y=1)}{\prod_{\alpha=1}^{d} P([x]_\alpha|y=0)}}\right)}$$

$$-\log \frac{\prod_{\alpha=1}^{d} P([x]_\alpha|y=1)}{\prod_{\alpha=1}^{d} P([x]_\alpha|y=0)} = -\sum_{\alpha=1}^{d} \frac{\frac{1}{\sqrt{2\pi}\,\theta_\alpha^2}\, e^{\frac{-(x-\mu_{0\alpha})^2}{2\theta_\alpha^2}}}{\frac{1}{\sqrt{2\pi}\,\theta_\alpha^2}\, e^{\frac{-(x-\mu_{1\alpha})^2}{2\theta_\alpha^2}}}$$

$$= -\sum_{\alpha=1}^{d} \log \frac{e^{\frac{-(x-\mu_{0\alpha})^2}{2\theta_\alpha^2}}}{e^{\frac{-(x-\mu_{1\alpha})^2}{2\theta_\alpha^2}}}$$

$$= -\sum_{\alpha=1}^{d} \frac{(x-\mu_{0\alpha})^2-(x-\mu_{1\alpha})^2}{2\theta_\alpha^2}$$

by simplifying the numerator:

$$= -\sum_{\alpha=1}^{d} \frac{x(\mu_{1\alpha}-\mu_{0\alpha})}{2\theta_\alpha^2} + \frac{\mu_{0\alpha}^2-\mu_{1\alpha}^2}{2\theta_\alpha^2}$$

Then, the entire equation becomes:

$$P(y=1|x) = \frac{1}{1+\exp\left(-\sum_{\alpha=1}^{d} \frac{x(\mu_{1\alpha}-\mu_{0\alpha})}{2\theta_\alpha^2} - \sum_{\alpha=1}^{d}\frac{\mu_{0\alpha}^2-\mu_{1\alpha}^2}{2\theta_\alpha^2} - \log\frac{P(y=1)}{P(y=0)}\right)}$$

$$P(y=1 \mid x) = \frac{1}{1 + \exp\left(-\sum_{d=1}^{d} \frac{x_d(\mu_{1d} - \mu_{0d})}{2\sigma_2^2} - w_0\right)}$$

$$w_0 = \sum \frac{\mu_{0d}^2 - \mu_{1d}^2}{2\sigma_2^2} + \log \frac{P(y=1)}{P(y=0)}$$

$$w_d = -\frac{\mu_{1d} - \mu_{0d}}{2\sigma_2^2}$$

$$P(y=1 \mid x) = \frac{1}{1 + \exp\left(\sum_{d=1}^{d} x_d w_d - w_0\right)}$$