CS446: Machine Learning, Fall 2017, Homework 3

Name: Lanxiao Bai (lbai5)

Worked individually

Problem 1

Solution:

$$\begin{aligned}
K(\mathbf{x},\mathbf{y}) &= (1+\mathbf{x}^T\mathbf{y})^2 \\
&= ([x_1,x_2,1]\cdot[y_1,y_2,1])^2 \\
&= (x_1y_1+x_2y_2+1)^2 \\
&= x_1^2y_1^2+2x_1y_1+2x_1x_2y_1y_2+x_2^2y_2^2+2x_2y_2+1 \\
&= [x_1^2,\sqrt{2}x_1,\sqrt{2}x_1x_2,x_2^2,\sqrt{2}x_2,1]\cdot[y_1^2,\sqrt{2}y_1,\sqrt{2}y_1y_2,y_2^2,\sqrt{2}y_2,1] \\
&= \phi(\mathbf{x})\phi(\mathbf{y})
\end{aligned}$$

As a result, we see that

$$\phi(\mathbf{x}) = [x_1^2,\sqrt{2}x_1,\sqrt{2}x_1x_2,x_2^2,\sqrt{2}x_2,1]$$

Problem 2

Solution: Since we have the objective function

$$L(w) = \frac{w^Tw}{2} + C\sum_{i\in[N]}\max(0,1-y_i\langle w,\phi(x_i)\rangle)$$

We can calculate the gradient

$$\begin{aligned}
\frac{\partial L(w)}{\partial w} &= 2w + C\sum_{i\in[N]}\frac{\partial\max(0,1-y_i\langle w,\phi(x_i)\rangle)}{\partial w} \\
&= 2w + C\sum_{i\in[N]}\begin{cases}0 & y_i\langle w,\phi(x_i)\rangle\geq 1\\ -y_i\frac{\partial\langle w,\phi(x_i)\rangle}{\partial w} & \textbf{otherwise}\end{cases} \\
&= 2w + C\sum_{i\in[N]}\begin{cases}0 & y_i\langle w,\phi(x_i)\rangle\geq 1\\ -y_i\phi(x_i) & \textbf{otherwise}\end{cases}
\end{aligned}$$

As a result, we have the SGD updating rule

$$w_{t+1} = w_t - \eta\frac{\partial L(w_t)}{\partial w_t} = w_t - 2\eta w_t + \begin{cases}0 & y_i\langle w,\phi(x_i)\rangle\geq 1\\ y_i\phi(x_i) & \textbf{otherwise}\end{cases}$$

References