Machine learning HW2

1. Recall that...

$$\min_{A,B} F(A,B) = \frac{1}{N}\sum_{n=1}^{N} \ln(1+\exp(-y_n(A\cdot(W_{svm}^T\phi(x_n)+b_{svm})+B)))$$

$$Z_n = W_{svm}^T\phi(x_n)+b_{svm}$$

$$P_n = \theta(-y_n(AZ_n+B)) \rightarrow \theta(s) = \frac{\exp(s)}{1+\exp(s)}$$

$$\nabla F(A,B) = ?$$

$\exp(\ \ ) \times (1+\exp(\ ))^{-1} \Rightarrow$

$$\frac{dF}{dA} \Rightarrow \frac{1}{N}\sum \ln(1+\exp(-y_n(AZ_n+B))) \Rightarrow$$

$$\Rightarrow \frac{1}{N}\sum \frac{1}{1+\exp(-y_n(AZ_n+B))} \times -y_n Z_n \times \exp(-y_n(AZ_n+B))$$

$$\Rightarrow \frac{1}{N}\sum -P_n \times y_n \times Z_n$$

$$\frac{dF}{dB} \Rightarrow \frac{1}{N}\sum \frac{1}{1+\exp(-y_n(AZ_n+B))} \times -y_n \times \exp(\ \ )$$

$$\Rightarrow \frac{1}{N}\sum -P_n \times y_n$$

$$A: \nabla F(A,B) = \frac{1}{N}\sum_{n=1}^{N}[-y_nP_nZ_n, -y_nP_n]^T$$

2. What is $H(F)$

$$\frac{\partial F^2}{\partial A^2} = -Y_n Z_n \frac{\partial P_n}{\partial A} \Rightarrow -Y_n Z \cdot (-YZ) P_n \times (1-P_n) \quad \underset{\underset{\Rightarrow 1}{\shortparallel}}{Y_n^2} Z_n^2 \cdot P_n \times (1-P_n)$$

$$\Rightarrow \underline{Z_n^2 \cdot P_n(1-P_n)}$$

$$\frac{\overbrace{\exp(-Y_n(AZ_n+B))}^{u}}{1+\underbrace{\exp(-Y_n(AZ_n+B))}_{u}}$$

$$\frac{dy}{du}\frac{du}{dA} \Rightarrow \underline{(1+u)^{-1}u} \Rightarrow -(1+u)^{-2} \Rightarrow \frac{1}{(1+u)^2} \times -Y_n Z \cdot u$$

$$\Rightarrow -YZ\frac{u}{(1+u)^2} \Rightarrow \frac{u}{(1+u)} \times \frac{1}{(1+u)} \Rightarrow \underline{P_n \times (1-P_n)}$$

$$\frac{\partial F}{\partial A \partial B} = -Y_n Z \frac{\partial P_n}{\partial B} \Rightarrow -\cancel{Y} Z \cdot \cancel{Y} \cdot P_n(1-P_n) \Rightarrow \underline{Z_n P_n(P_n-1)}$$

$$\frac{\partial P_n}{du} \Rightarrow \frac{u}{1+u} \Rightarrow u(1+u)^{-1} \Rightarrow u \times -(1+u)^{-2} + (1+u)^{-1}$$

$$\Rightarrow \frac{-u}{(1+u)^2} + \frac{1+u}{(1+u)^2} \Rightarrow \frac{1}{(1+u)^2}$$

$$\frac{du}{dB} \Rightarrow \underline{\underline{-Y_n \cdot u}}$$

$$\frac{dy}{du} \times \frac{du}{dB} \Rightarrow -Y\frac{u}{(1+u)^{-2}} \Rightarrow \underline{\underline{-Y \times P_n \times (1-P_n)}}$$

$$-\cancel{Y} \times Z \times -\cancel{Y} \times P_n(1-P_n) \Rightarrow Z \times P_n(1-P_n)$$

$$\frac{\partial F}{\partial B \partial A} = -Y\frac{\partial P}{\partial A} \Rightarrow -Y \times P_n(1-P_n)$$

$$A: \begin{bmatrix} \frac{\partial^2 F}{\partial A^2} & \frac{\partial^2 F}{\partial A \partial B} \\ \frac{\partial^2 F}{\partial B \partial A} & \frac{\partial^2 F}{\partial B^2} \end{bmatrix}$$

$$\frac{\partial F}{\partial B \partial B} = -Y \times \frac{\partial P}{\partial B} \Rightarrow -Y \times -Y \cdot P_n(1-P_n) \Rightarrow P_n(1-P_n)$$

3.

kernel ridge regression

$B = (\lambda \underset{N \times N}{I} + K)^{-1} Y$

4.

$(P_2)\ \min_{b, w, \xi^\vee, \xi^\wedge} \frac{1}{2} w^T w + C \sum_{n=1}^{N} \left((\xi_n^\vee)^2 + (\xi^\wedge)^2 \right)$

such that $-\epsilon - \xi^\vee \le y_n - w^T \phi(x_n) - b \le \epsilon + \xi_n^\wedge$

Find unconstrained form

Standard SVM Regression Primal

$\min \frac{1}{2} w^T w + C \sum_{n=1}^{N} \max(0, |w^T z_n + b - y| - \epsilon)$

5. $\underline{s_n = \sum_{m=1}^{N} \beta_m K(x_n, x_m) + b}$

w_* for (P_2) in Question 4 must satisfy $\underline{w_* = \sum_{n=1}^{N} \beta_n z_n}$

$\min_{b, \beta} F(b, \beta) = \frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} \beta_n \beta_m K(x_n, x_m) +$ something

$C \sum_{n=1}^{N} (\max(0, |y - \underline{w}^T \phi(x_n) - b| - \epsilon)$

$\max(0, |y - w^* \phi(x_n - b| - \epsilon)$

$\Downarrow$

$|y_n - s_n| - \epsilon$

$\Downarrow$

if $\underline{|y_n - s_n| \ge \epsilon}$ then do $\max(0, \ )$

$(|y - s_n| - \epsilon)^2 \to (|y - s_n| - \epsilon) - 2K(x_n, x_m)$

$\Rightarrow -2C \sum [[|y_n - s_n| \ge \epsilon]] (|y_n - s_n - \epsilon) K(x_n, x_m)$