$$F(A,B) = \frac{1}{N}\sum \ln\left(1+\exp\left(-y_n(A\cdot z_n + B)\right)\right)$$

1. $$\frac{\partial F(A,B)}{\partial A} = \frac{1}{N}\sum^{N} \frac{\exp(-y_n(Az_n+B))}{1+\exp(-y_n(Az_n+B))}\cdot(-y_n z_n) = -\frac{1}{N}\sum_{n=1}^{N} p_n y_n z_n$$

$$\frac{\partial F(A,B)}{\partial B} = -\frac{1}{N}\sum_{n=1}^{N} p_n y_n$$

$$\nabla F(A,B) = -\frac{1}{N}\sum_{n=1}^{N} y_n p_n (z_n, 1)$$

2. $$H(F) = \begin{pmatrix} \frac{\partial^2 F}{\partial A^2} & \frac{\partial^2 F}{\partial A \partial B} \\ \frac{\partial^2 F}{\partial B \partial A} & \frac{\partial^2 F}{\partial B^2} \end{pmatrix}$$

$$\frac{\partial^2 F(A,B)}{\partial A^2} = \frac{\partial}{\partial A}\left(-\frac{1}{N}\sum p_n y_n z_n\right) = \frac{1}{N}\sum_{n=1}^{N} z_n^2 p_n(1-p_n)$$

$$\frac{\partial^2 F(A,B)}{\partial A \partial B} = \frac{1}{N}\sum^{N} z_n p_n(1-p_n)$$

$$\frac{\partial^2 F(A,B)}{\partial B^2} = \frac{1}{N}\sum p_n(1-p_n)$$

$$H(F) = \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)\begin{bmatrix} z_n^2 & z_n \\ z_n & 1 \end{bmatrix}$$

3. when $\gamma \to \infty$ $\quad K(x,x') = \exp(-\gamma\|x-x'\|^2) \to 0$
$(x \neq x')$

$\Rightarrow$ $b = 1$ or -1. $\quad \alpha = 0$.

4. two samples $\begin{cases} (x_1, & x_1 - x_1^2) \\ (x_2, & \underbrace{x_2 - x_2^2}_{y}) \end{cases}$

$$E_{ols} = (w_1 x_1 + w_0 - x_1 + x_1^2)^2 + (w_1 x_2 + w_0 - x_2 + x_2^2)^2$$

$$\frac{\partial E_{ols}}{\partial w_1} = 2(w_1 x_1 + w_0 - x_1 + x_1^2) x_1 + 2(w_1 x_0 + w_0 - x_2 + x_2^2) x_2 = 0$$

$$\Rightarrow \quad w_1 = (1 - x_1 - x_2)$$

$$\frac{\partial E_{ols}}{\partial w_0} = 2(w_1 x_1 + w_0 - x_1 + x_1^2) + 2(w_1 x_2 + w_0 - x_2 + x_2^2) = 0$$

$$\Rightarrow \quad w_0 = +x_1 x_2$$

$$\Rightarrow \quad g(x) = (1 - x_1 - x_2)\, x + x_1 \cdot x_2$$

$$\bar{g}(x) = \mathcal{E}[g(x)] = \frac{1}{4}$$

5. $$u_n (y_n - w^T x_n)^2 = (\sqrt{u_n}\, y_n - w^T \sqrt{u_n}\, x_n)^2$$

$$\{(\tilde{x}_n, \hat{y}_n)\}_{n=1}^{N} = \{(\sqrt{u_n}\, x_n, \sqrt{u_n}\, y_n)\}_{n=1}^{N}$$

6. error rate $= 0.22 = \varepsilon_t$

$$\frac{u_+^{(2)}}{u_-^{(2)}} = \frac{\varepsilon_t}{1 - \varepsilon_t} = \frac{22}{78} = \frac{11}{39}$$

7. First, all positive and negative are two decision stumps (D-S)

For each dimension, there are $2 \times M$ interval between $-M$ and M, each could be $+1$ or -1

total d dimensions, $\Rightarrow$ total $4dM + 2$ DSs.

8. $K_{ds}(x, x')$ is inner product of $\phi_{ds}(x)$

$$= 4dM + 2 - 2 \cdot \sum_{n=1}^{|q|} |x_n - x_n'|$$

if $x = x'$ $\Rightarrow K(x, x') = 4dM + 2$.

Double A