Homework 1

Theory

Tianyang Chen

Part A Theory

1. Maximum Posterior vs Probability of Chance

If $P(w_{max}|\vec{x}) < \frac{1}{c}$

Then $\sum_{i=1}^{c} P(w_i|\vec{x}) < \frac{1}{c} \cdot c = 1$

$\therefore P(w_{max}|\vec{x}) \geq \frac{1}{c}$

$P(error) = 1 - P(w_{max}|\vec{x}) \leq 1 - \frac{1}{c} = \frac{c-1}{c}$

2. Bayes Decision Rule Classifier

choose w_1 if $P(w_1)P(\vec{x}|w_1) > P(w_2)P(\vec{x}|w_2)$

$\because P_{i1} = P(x_i=1|w_1) = P$, $P(x_i=0|w_1) = 1-P$

$P_{i2} = P(x_i=1|w_2) = 1-P$, $P(x_i=0|w_2) = 1-P$

$\therefore P(w_1)P^k(1-P)^{d-k} > P(w_2)\cdot(1-P)^k \cdot P^{d-k}$, $k \in [1, d]$ and is an integer

$\Rightarrow P^{2k-d} > (1-P)^{2k-d}$

$\Rightarrow 2k-d > 0$

$\Rightarrow k > \frac{d}{2}$

$\Rightarrow$ choose w_1 if $\sum_{i=1}^{d} x_i > \frac{d}{2}$

3. The Ditzler Household Growing Up

BB — denotes two boys

Greg — denotes a boy name Greg

$$P(BB|Greg) = \frac{P(BB, Greg)}{P(Greg)} = \frac{\frac{2}{8}}{\frac{4}{8}} = \frac{1}{2}$$

4. Linear classifier with a Margin

$$\begin{cases} w^T x_1 + b = 1 \\ w^T x_2 + b = -1 \end{cases}, \quad \arg\min_{w \in R^p} \|w\|_2^2$$

$$L = \|w\|_2^2 + \lambda_1(w^T x_1 + b - 1) + \lambda_2(w^T x_2 + b + 1)$$

$$\begin{cases} \frac{\partial L}{\partial w} = 2w + \cancel{\lambda_1 w + \lambda_2 w} \ \lambda_1 x_1 + \lambda_2 x_2 = 0 \\ \frac{\partial L}{\partial b} = \lambda_1 + \lambda_2 = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \lambda_1 = -\lambda_2 \\ \vec{w} = \frac{1}{2}\lambda_1(\vec{x}_2 - \vec{x}_1) \\ b = -\frac{1}{2}\vec{w}^T(\vec{x}_1 + \vec{x}_2) \end{cases}$$

5. Decision Making with Bayes

$P(w|x)$ is the probability we are interested in but we can not get this probability directly. So we use Bayes rule to compute $P(w|x)$. However, we get $P(x|w)$ and $P(w)$ by statistical data, it's not 100% accurate, which causes the inaccuracy of $P(w|x)$.

If we know $P(x)$, we can estimate the density of event, ~~we~~ for those $\vec{x}$ with high probability, we can try to do subclassification, in order to get better prediction.