Homework 2

2. (1) $H(t) = -\sum_c p(c|t)\log_2 p(c|t) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = 1$

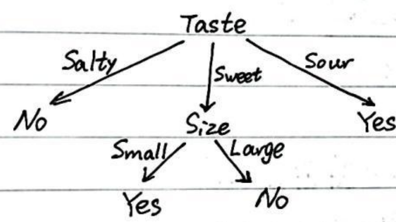
(2) $InfoGain_{split} = H(t) - \sum_{k=1}^{3}\frac{n_k}{n}H(k)$

$= 1 - \left(\frac{3}{10}\times(-1\cdot\log_2 1) + \frac{4}{10}\times(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}) + \frac{3}{10}\times(-1\cdot\log_2 1)\right)$

$= 1 - \frac{2}{5}$

$= \frac{3}{5}$

(3)




3. (1) $L(\theta) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$

$\ln L(\theta) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

We need $\ln L(\theta)$ ~~to~~ differentiate to $\mu$ and $\sigma^2$ and equal to 0.

$$\frac{\partial \ln L(\theta)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \Rightarrow \sum_{i=1}^{n}x_i - n\mu = 0 \Rightarrow \hat{\mu} = \bar{x}$$

$$\frac{\partial \ln L(\theta)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

(2) $E(\hat{\mu}) = E\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right) = \frac{1}{n}\sum_{i=1}^{n}E(x_i) = \frac{1}{n}\cdot n\cdot\mu = \mu$

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n}x_j\right)^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j$

Suppose $\mu - x_i = \Delta_i$

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\mu-\Delta_i)^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\mu-\Delta_i)(\mu-\Delta_j)$

$= \mu^2 - \frac{2\mu}{n}\sum_{i=1}^{n}\Delta_i + \frac{1}{n}\sum_{i=1}^{n}\Delta_i^2 - \mu^2 + \frac{\mu}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\Delta_i+\Delta_j) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i\Delta_j$

$E(\Delta_i) = 0$, $E(\Delta_i^2) = \sigma^2$, $E(\Delta_i\Delta_j) = 0$

$E(\hat{\sigma}^2) = -\frac{2\mu}{n}\cdot 0 + \frac{1}{n}\sum_{i=1}^{n}\sigma^2 + \frac{\mu}{n^2}\cdot 0 + \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n-1}{n}\sigma^2$

Hence, $E\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \sigma^2$, the conclusion holds.

No.
Date

4. ① Suppose $P(y_i) = \prod_{k=1}^{c} p_k^{I(y_i=k)}$

Likelihood function $L(p_1, p_2, \cdots, p_c) = \prod_{i=1}^{n} P(y_i) = \prod_{i=1}^{n}\prod_{k=1}^{c} p_k^{I(y_i=k)}$

$\ln L = \sum_{i=1}^{n}\sum_{k=1}^{n} I(y_i=k)\ln(p_k)$

By Lagrangian multiplier method,

$l(p_1, \cdots, p_c, \lambda) = \sum_{i=1}^{n}\sum_{k=1}^{c} I(y_i=k)\ln p_k - \lambda(\sum_{k=1}^{c} p_k - 1)$

To get MLE estimator of $p_k$, $\frac{\partial l}{\partial p_k} = 0$.

$\frac{\partial l}{\partial p_k} = \sum_{i=1}^{n}\frac{I(y_i=k)}{p_k} - \lambda = 0 \Rightarrow \hat{p}_k = \frac{\sum_{i=1}^{n} I(y_i=k)}{\lambda}$

Set $\lambda = n$, then $\hat{p}_k = \frac{1}{n}\sum_{i=1}^{n} I(y_i=k)$

② Suppose $P(x_i|y_i) = \prod_{s=1}^{t}\prod_{k=1}^{c} p_{sk}^{I(x_i=s, y_i=k)}$

Then $L(p_{11}, \cdots, p_{tc}) = \prod_{i=1}^{n} p(x_i|y_i) = \prod_{i=1}^{n}\prod_{s=1}^{t}\prod_{k=1}^{c} p_{sk}^{I(x_i=s, y_i=k)}$

$l(p_{11}, \cdots, p_{tc}) = \ln L = \sum_{i=1}^{n}\sum_{s=1}^{t}\sum_{k=1}^{c} I(x_i=s, y_i=k)\ln p_{sk}$

By Lagrangian multiplier method,

$l(p_{11}, \cdots, p_{tc}, \lambda) = \sum_{i=1}^{n}\sum_{s=1}^{t}\sum_{k=1}^{c} I(x_i=s, y_i=k)\ln p_{sk} - \lambda(\sum_{s=1}^{t} p_{sk} - 1)$

$\frac{\partial l}{\partial p_{sk}} = \sum_{i=1}^{n}\frac{I(x_i=s, y_i=k)}{p_{sk}} - \lambda = 0 \Rightarrow \hat{p}_{sk} = \frac{1}{\lambda}\sum_{i=1}^{n} I(x_i=s, y_i=k)$

According to $\sum_{s=1}^{t} p_{sk} = 1$, set $\lambda = \sum_{i=1}^{n} I(y_i=k)$

then $\hat{p}_{sk} = \frac{\sum_{i=1}^{n} I(x_i=s, y_i=k)}{\sum_{i=1}^{n} I(y_i=k)}$

5. We can first sample $n$ unlabeled examples, $S_x = (x_1, \cdots, x_n)$, according to $P_x$, and an additional unlabeled example, $x \sim P_x$, then find $\pi_s(x)$ to be the nearest neighbour of $x$ in $S_x$, and finally sample $y \sim \eta(x)$ and $y_{\pi_s(x)} \sim \eta(\pi_s(x))$. Then

$$\underset{S}{E}(\mathcal{E}(f^{1NN})) = \underset{S_x \sim P_x^n,\, x\in P,\, y\sim\eta(x),\, y'\sim\eta(\pi_s(x))}{E}(1_{(y\neq y')})$$

$$= \underset{S_x \sim P_x^n,\, x\in P}{E}\left(\underset{y\sim\eta(x),\, y'\sim\eta(\pi_s(x))}{P}(y\neq y')\right)$$

$$\underset{y\sim\eta(x),\, y'\sim\eta(\underbrace{\pi_s(x)}_{\text{Suppose is } x'})}{P}(y\neq y') = \eta(x')(1-\eta(x)) + (1-\eta(x'))\eta(x)$$

$$= 2\eta(x)(1-\eta(x)) + (\eta(x)-\eta(x'))(2\eta(x)-1)$$

$\eta$ is $c$-Lipschitz, $|\eta(x)-\eta(x')| \le c\|x-x'\|$, and $|2\eta(x)-1| \le 1$

$\underset{y\sim\eta(x),\, y'\sim\eta(x')}{P}(y \neq y') \le 2\eta(x)(1-\eta(x)) + c\|x-x'\|$

Hence, $\underset{S}{E}[\mathcal{E}(f^{1NN})] \le \underset{x}{E}[2\eta(x)(1-\eta(x))] + c\underset{S,x}{E}[\|x-x'\|]$

$\underset{x}{E}[\eta(x)(1-\eta(x))] \le \underset{x}{E}[\min\{\eta(x), 1-\eta(x)\}] = \mathcal{E}(f^*)$

Then, $\underset{S}{E}[\mathcal{E}(f^{1NN})] \le \mathcal{E}(f^*) + c\underset{S\sim P^n}{E}\ \underset{x\sim P_x}{E}[\|x - x_{\pi_S(x)}\|]$