ComS 573 Machine Learning – HW2 solution

Problem 1:

(20 pts.) Let $\vec{x} = (x_1, \ldots, x_d)^t$ be a d-dimensional binary (0 or 1) vector with a multivariate Bernoulli distribution

$$P(\vec{x}|\vec{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i},$$

where $\vec{\theta} = (\theta_1, \ldots, \theta_d)^t$ is an unknown parameter vector, θ_i being the probability that $x_i = 1$. Given i.i.d. data set $D = \{\vec{x}_1, \ldots, \vec{x}_n\}$, *derive* the maximum-likelihood estimate for $\vec{\theta}$.

Solution:

Likelihood of $\vec{\Theta}$ w.r.t. observed samples:

$$P(D|\vec{\Theta}) = \prod_{k=1}^{n} P(\overrightarrow{x_k}|\vec{\Theta})$$

$$= \prod_{k=1}^{n}\prod_{i=1}^{d} \Theta_i^{x_{ki}} \ (1-\Theta_i)^{1-x_{ki}}$$

Where $x_{ki} = \begin{cases} 1 & \text{if i appears in kth sample of D} \\ 0 & \text{otherwise} \end{cases}$

The log-likelihood function is given as follows:

$$LL(\vec{\Theta}) = ln\, P(D|\vec{\Theta})$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{d} ln\,(\Theta_i^{x_{ki}} \ (1-\Theta_i)^{1-x_{ki}})$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{d} (ln\, \Theta_i^{x_{ki}} + ln\,(1-\Theta_i)^{1-x_{ki}})$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{d} (x_{ki}\, ln\, \Theta_i + (1\text{-}x_{ki})\, ln\,(1-\Theta_i)) \qquad \ldots\ldots\ldots\ldots\ldots eq(1)$$

Optimal estimation:

To determine the $\vec{\Theta}$ that maximizes the log-likelihood, take partial darivative of ln $P(D|\vec{\Theta})$ with respect to Θ_i and set to zero.

$$\frac{\partial\, ln\, P(D|\vec{\Theta})\,)}{\partial \Theta_i} = \sum_{k=1}^{n}(\frac{x_{ki}}{\Theta_i} - \frac{(1\text{-}x_{ki})}{(1-\Theta_i)}) = 0$$

$$\sum_{k=1}^{n}(\frac{x_{ki}}{\Theta_i} \quad - \sum_{k=1}^{n}\frac{(1-x_{ki})}{(1-\Theta_i)} \quad = 0$$

$$\frac{N_i^1}{\Theta_i} - \frac{(n - N_i^1)}{(1 - \Theta_i)} = 0$$

$$\text{Where } N_i^1 = \sum_{k=1}^{n} x_{ki} \qquad (\text{ the \# of samples}, x_{ki} = 1 \text{ in D.})$$

$$\text{And } (n - N_i^1) = \sum_{k=1}^{n}(1 - x_{ki}) \qquad (\text{ the \# of samples}, x_{ki} = 0 \text{ in D.})$$

Therefore,

$$\Theta_{i\text{MLE}} = \frac{N_i^1}{n}$$

$$\vec{\Theta}_{\text{MLE}} = (\frac{N_1^1}{n}, \frac{N_2^1}{n}, \ldots\ldots\ldots\ldots\ldots\ldots \frac{N_i^1}{n} \ldots\ldots\ldots\ldots \frac{N_d^1}{n})$$

Problem 2:

(10 pts.) Assume that a classifier correctly classifies 900 of the 1000 examples in the test set. What is the estimated accuracy of the classifier? Give 95% confidence interval.

Solution:

A classifier correctly classifies 900 of the 1000 examples in the test set. Thus, the mean value of estimated accuracy of the classifier is as follows:

$$\mu = 900/1000 = 0.9$$

standard deviation can

$$\sigma = \sqrt{0.9(1-0.9)/1000}$$

For 95% confidence interval, $Z_N = 1.96$ (looking up the table)

Confidence interval lies in the interval [$\mu \pm Z_N . \sigma$], therefore, 95% confidence interval is [0.881, 0.909].