MA2823: Introduction to Machine Learning **Homework 4**

Due October 20, 2017.

Turn in your homework

- as a PDF file (**This means no .docx**)
- named HW04_<LastName>_<FirstName>.pdf (no accents, no spaces, no < or > characters)
- at http://tinyurl.com/ma2823-2017-hw

Question 1

Assume p random variables $X_1, \dots, X_p$, conditionally independent given Y. Y is a discrete random variable that can take one of K values $\{y_1, \dots, y_K\}$, corresponding to K classes.

We suppose that each X_j is distributed normally:

$$p(X_j = u | Y = y_k) = \frac{1}{\sigma_{jk}\sqrt{2\pi}} \exp\left(\frac{-(u-\mu_{jk})^2}{2\sigma_{jk}^2}\right).$$

Let us assume that the variance is independent of the class k and the feature j, i.e. $\sigma_{jk} = \sigma$.

We observe n datapoints $\{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\}$ and their labels $\{y^1, \dots, y^n\}$. Derive the maximum likelihood estimator for μ_{jk}.

To simplify notations, you can use the indicator $I_{ik} = \begin{cases} 1 & \text{if } y^i = y_k \\ 0 & \text{otherwise.} \end{cases}$

Solution: The likelihood of the parameters is given by

$$\begin{aligned}\mathcal{L}(\mu_{jk}, \sigma) &= \prod_{i=1}^{n} p(X_j = x_j^i | \mu_{jk}, \sigma)^{I_{ik}} \\ &= \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x_j^i - \mu_{jk})^2}{2\sigma^2}\right)\right)^{I_{ik}}.\end{aligned}$$

The log-likelihood is:

$$l(\mu_{jk}, \sigma) = \sum_{i=1}^{n} I_{ik} \left[\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \left(\frac{-(x_j^i - \mu_{jk})^2}{2\sigma^2}\right)\right].$$

Taking the derivative with respect to μ_{jk} and setting it to 0:

$$\frac{\partial l(\mu_{jk}, \sigma)}{\partial \mu_{jk}} = 0,$$

we obtain

$$\hat{\mu}_{jk} = \frac{\sum_{i=1}^{n} I_{ik} x_j^i}{\sum_{i=1}^{n} I_{ik}}.$$

Note that $\sum_{i=1}^{n} I_{ik} x_j^i$ is the empirical mean of samples from class k and $\sum_{i=1}^{n} I_{ik}$ is the number of such samples.