Course: Machine learning

By: Pavlo Mozharovskyi

Tutorial to Lecture 1: Brief introduction to machine learning

Task 1: Explore the Fisher's linear discriminant analysis (LDA).

(a) Demonstrate on a short example that $\hat{R}_1$ (see Appendix) is a consistent estimator (for class "1") of the LDA error rate.

(b) Obtain empirical distribution of the LDA error rate and of its estimator as $\hat{R}_1$ by calculating these over 2 000 random draws.

(c) Compare these two distributions and verify their normality visually (use `R`-functions `density` and `dnorm`), with Shapiro-Wilk test (use `R`-function `shapiro.test`), and by the means of the QQ-plot (use `R`-function `qqnorm`).

Perform experiments in part (a) with training sample sizes 10 000 or (and) 100 000; in part (b) with training sample size 1000.

Perform (a), (b), and (c) for three cases of normal distribution, namely:

$$\text{Model 1: } X|Y=0 \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 1\\1 & 4\end{pmatrix}\right) \text{ and } X|Y=1 \sim N\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1 & 1\\1 & 4\end{pmatrix}\right);$$

$$\text{Model 2: } X|Y=0 \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 1.1\\1.1 & 3\end{pmatrix}\right) \text{ and } X|Y=1 \sim N\left(\begin{pmatrix}1.4\\2.1\end{pmatrix}, \begin{pmatrix}1 & 1.1\\1.1 & 3\end{pmatrix}\right);$$

$$\text{Model 3: } X|Y=0 \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 2.1\\2.1 & 9\end{pmatrix}\right) \text{ and } X|Y=1 \sim N\left(\begin{pmatrix}3\\2.5\end{pmatrix}, \begin{pmatrix}1 & 2.1\\2.1 & 9\end{pmatrix}\right).$$

Interpret the results.

Task 2: Find optimal k for the k-nearest neighbors classifier (kNN).

For models 1–3 from Task 1, study the behavior of the error rate of kNN dependent on k — the number of nearest neighbors. For this, plot error rate estimated by leave-one-out cross-validation (jackknife) against k. Increase sample size until having noticed a tendency, consider (odd) ks between 1 and half of the sample size. Use `R`-functions `knn` and `knn.cv` from `R`-package `class`.

Interpret the results.

Appendix

Error probability R_1 can be consistently estimated:

$$\hat{R}_1 = \Phi\big(\frac{\hat{u}_0}{\sqrt{\hat{v}_0}}\big) \,,$$

where

$$\begin{aligned}
\hat{u}_0 &= -\frac{\hat{\Delta}^2}{2(1-\frac{d}{n})} \,, \\
\hat{v}_0 &= \frac{1}{(1-\frac{d}{n})^3}\big(\hat{\Delta}^2 + \frac{d}{n\pi_0\pi_1}\big) \,, \\
\hat{\Delta}^2 &= \frac{n-d-1}{n}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_0)^T \boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_0) - \frac{(n+2)d}{n_0 n_1} \,.
\end{aligned}$$

with n being the length of the training sample, d — dimension, π_0 and π_1 — prior probabilities of the corresponding classes, n_0 and n_1 — their sample sizes, $\bar{\boldsymbol{x}}_0$, $\bar{\boldsymbol{x}}_1$, and $\boldsymbol{S}$ — consistent mean and covariance estimates.