Pattern Analysis and Machine Intelligence

Matteo Matteucci, Davide Eynard

06/07/2015

1 Statistical learning (8 points)

(a) Both classification and regression problems can be solved by simple but restrictive methods as well as more complex but flexible ones. Explain which ones are better:

1. when doing *inference* or *prediction*;
2. when the irreducible error is extremely high;
3. when the number of observations is very large and the number of predictors is small;
4. when the function we need to estimate is highly non-linear.

(b) What are the *Bayes classifier* and the *Bayes error rate*? Define them and explain why the latter is defined as analogous to the irreducible error.

2 Linear regression (8 points)

Given the variables $x_1 = \{14, 1, 13, 8, 11, 19, 0, 9\}$, $x_2 = \{12, 13, 7, 10, 8, 11, 16, 10\}$, and $y = \{26, 7, 24, 15, 24, 38, 2, 13\}$, manually calculate the parameters $\hat{\beta_0}$ and $\hat{\beta_1}$ of the two linear functions $\hat{y} = \hat{\beta_0} + \hat{\beta_1} x_1$ and $\hat{y} = \hat{\beta_0} + \hat{\beta_1} x_2$ which fit the given data. To ease your calculations, take the following steps:

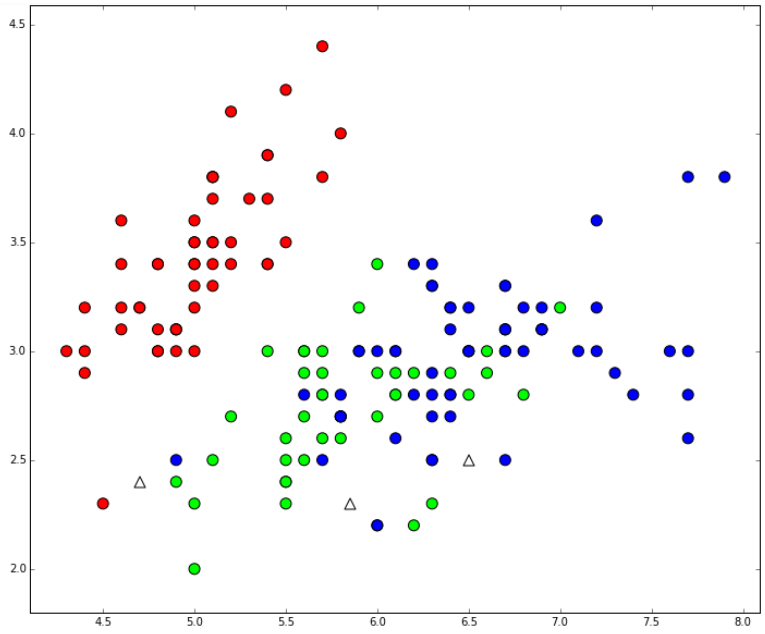
1. calculate the mean $\bar{x}$ of x and the mean $\bar{y}$ of y
2. calculate x-$\bar{x}$ (a vector where each value is $x_i - \bar{x}$) and y-$\bar{y}$
3. calculate $\hat{\beta_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, then $\hat{\beta_0} = \bar{y} - \beta_1 \hat{x}$

Measure the quality of your predictions in terms of MSE, compare the results of the two linear regressions, and justify them.

3 Classification (8 points)

(a) Given the subset of Iris dataset in figure, classify the three points identified with white triangles (at coordinates $(5.85, 2.3)$, $(4.7, 2.4)$, and $(6.5, 2.5)$ respectively), using the KNN algorithm with $k = 3, 5, 7$. Note: if your point has the same amount of neighbors for each class, you can assign it the class of the closest one.

(b) Explain what the "curse of dimensionality" is. How would you address this problem?



4 Clustering (8 points)

Given the following algorithms:

1. K-means
2. Hierarchical
3. Mixture of Gaussians
4. DBSCAN
5. K-medoids
6. Fuzzy C-means
7. Jarvis-Patrick

complete the following sentences matching them with one (or more!) of the algorithms, answering the questions in parentheses and providing detailed explanations to motivate your choices. (NOTE: although sentences refer to a single algorithm, there may be more than one valid choice. In these cases, provide and motivate all of them).

- (a) This algorithm relies on a "self-scaling" neighborhood (what does this mean? How can this be accomplished?)
- (b) This algorithm builds new clusters by merging or splitting existing ones (describe the differences between the two approaches and provide the computational complexity of this approach)
- (c) This algorithm provides a "soft" classification (what does this mean?)
- (d) This algorithm can provide good results even if noise is present in the dataset (is it also able to detect which points are noise?)