DATA SCIENCE & MACHINE LEARNING (part - 3)

techworldthink • March 06, 2022

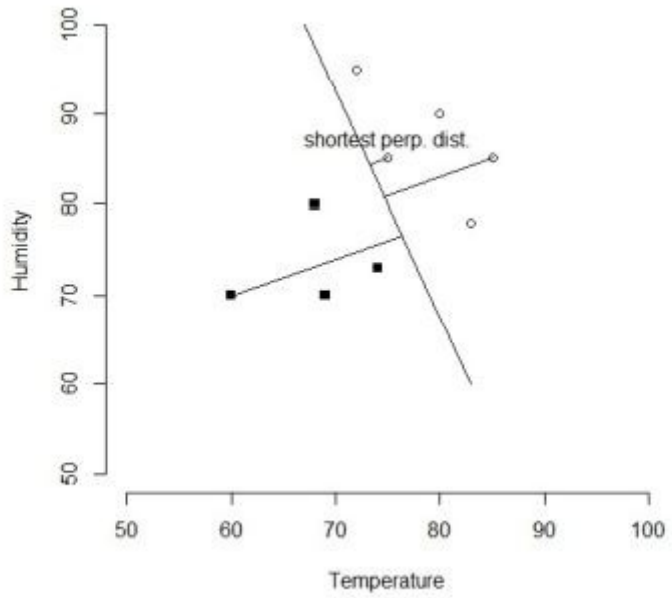
8. What is maximum margin hyperplane.

Margin of a separating line

To choose the “best” separating line, we introduce the concept of the margin of a separating line. Given a separating line for the data, we consider the perpendicular distances of the data points from the separating line. The double of the shortest perpendicular distance is called the “margin of the separating line”.

Maximal margin separating line

The “best” separating line is the one with the maximum margin



The separating line with the maximum margin is called the “maximum margin line” or the “optimal separating line”. This line is also called the “support vector machine”.

9. Define precision, recall and F-measure.

In machine learning, precision and recall are two measures used to assess the quality of results produced by a binary classifier. They are formally defined as follows.

Let a binary classifier classify a collection of test data. Let

TP = Number of true positives

TN = Number of true negatives

FP = Number of false positives

FN = Number of false negatives

The precision P is defined as P = TP / (TP + FP)

The recall R is defined as R = TP / (TP + FN)

F-measure = (2 × TP) / (2 × TP + FP + FN)

The F-measure of the system is defined as **the weighted harmonic mean of its precision and recall.**

$$F_1 = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})}$$

10. Explain bootstrap sampling

> *In statistics, Bootstrap Sampling is a method that involves drawing of sample data repeatedly with replacement from a data source to estimate a population parameter.*

- Sampling: With respect to statistics, sampling is the process of selecting a subset of items from a vast collection of items (population) to estimate a certain characteristic of the entire population
- Sampling with replacement: It means a data point in a drawn sample can reappear in future drawn samples as well

- Parameter estimation: It is a method of estimating parameters for the population using samples. A parameter is a measurable characteristic associated with a population. For example, the average height of residents in a city, the count of red blood cells, etc.

1. It is the building block for many modern machine learning algorithms
2. It can be used to estimate the parameters of a population