Parametric vs Non-parametric Models

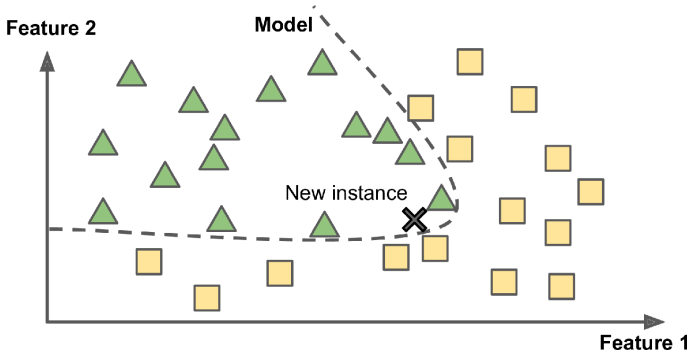
K-nearest neighbors

Jose Luis Paniagua Jaramillo
jlpaniagua@uao.edu.co

Outline

Parametric vs Non-parametric Models I

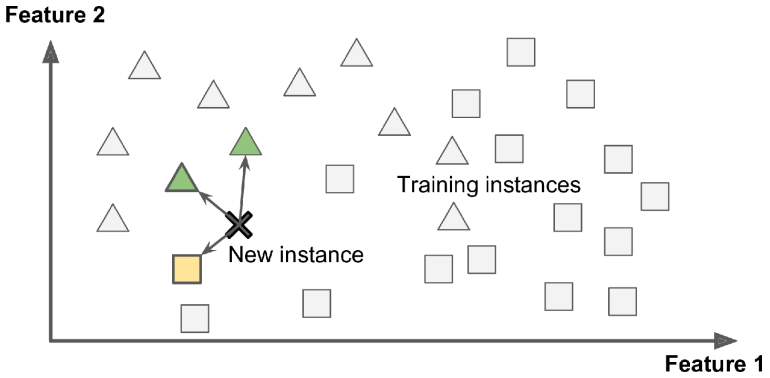
Model-based learning.



we estimate parameters from the training dataset to learn a function that can classify new data points without requiring the original training dataset anymore [1].

Parametric vs Non-parametric Models II

Instance-based learning.



the number of parameters changes with the amount of training data [1].

k-nearest neighbors I

1. Choose the number of k and a distance metric
2. Find the k-nearest neighbors of the data record that we want to classify
3. Assign the class label by majority vote

Distance Metric

$$d(x^{(i)}, x^{(j)}) = \sqrt[p]{\sum_k \left| x_k^{(i)} - x_k^{(j)} \right|^p}$$

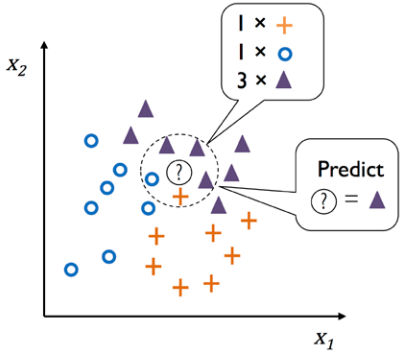


Figura: [2]

Ejemplo

Implement a knn algorithm for the iris dataset:

1. Use Petal length and Petal width (do not normalize the data).
2. Use Petal length and Petal width (normalize the data).
3. Use all features

Analyze the results.

References

[1] Brad Boehmke y Brandon Greenwell. *Hands-on machine learning with R*. Chapman y Hall/CRC, 2019.

[2] Sebastian Raschka, Joshua Patterson y Corey Nolet. “Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence”. En: *Information* 11.4 (2020), pág. 193.