Sapienza University of Rome

Master in Engineering in Computer Science

Machine Learning

A.Y. 2022/2023

Prof. Luca Iocchi

10. Instance based learning

Luca Iocchi

Summary

- Non-parametric models
- K-NN for classification
- Locally weighted regression

References

C. Bishop. Pattern Recognition and Machine Learning. Sect. 2.5

Parametric and non-parametric models

Parametric model: Model has a fixed number of parameters

Examples:

- Linear regression
- Logistic regression
- Perceptron
- ...

Non-parametric model: Number of parameters grows with amount of data

Simple non-parametric model: **instance-based learning**

K-nearest neighbors

Classification problem: $f : X \mapsto C$ with data set $D = \{(\mathrm{x}_n, t_n)_{n=1}^N\}$

Classification with K-NN,

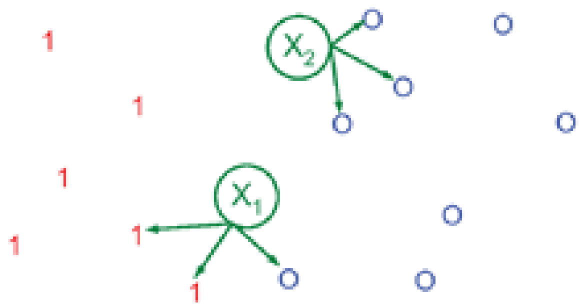
1. Find K nearest neighbors of new instance x
2. Assign to x the most common label among the majority of neighbors

Likelihood of class c for new instance x:

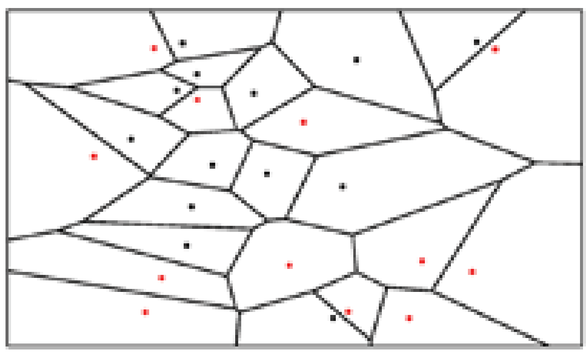
$$p(c|\mathrm{x}, D, K) = \frac{1}{K} \sum_{\mathrm{x}_n \in N_K(\mathrm{x}_n, D)} \mathbb{I}(t_n = c),$$

with $N_K(\mathrm{x}_n, D)$ the K nearest points to x_n and $\mathbb{I}(e) = \begin{cases} 1 & \text{if } e \text{ is true} \\ 0 & \text{if } e \text{ is false} \end{cases}$.

K-nearest neighbors examples



$K = 3$



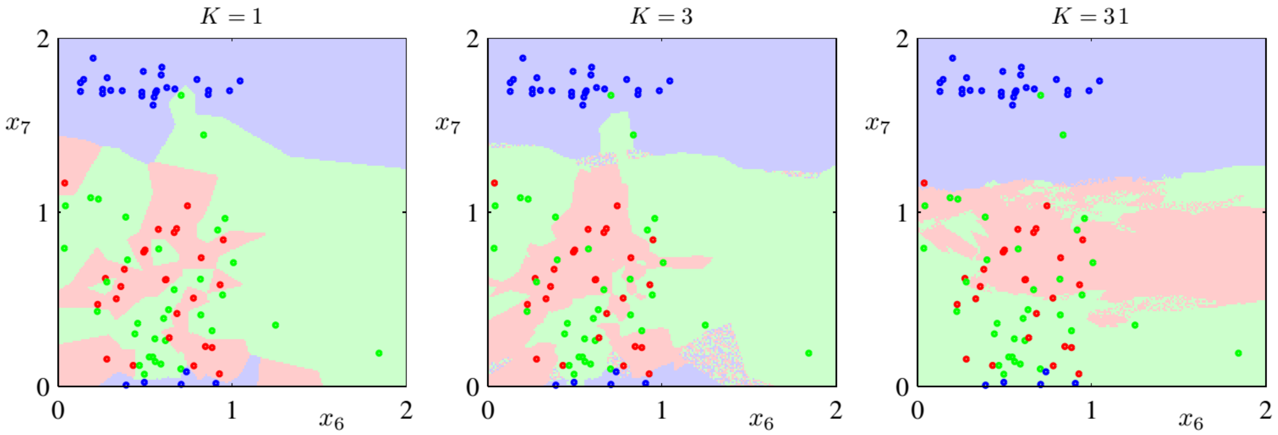
Voronoi tesselation for $K = 1$

Requires storage of all the data set!

Depends on a distance function!

K-nearest neighbors

Increasing K brings to smoother regions (reducing overfitting)



Kernelized nearest neighbors

Distance function in computing $N_K(\mathrm{x}, D)$

$$\|\mathrm{x} - \mathrm{x}_n\|^2 = \mathrm{x}^T\mathrm{x} + \mathrm{x}_n^T\mathrm{x}_n - 2\mathrm{x}^T\mathrm{x}_n.$$

can be kernelized by using a kernel $k(\mathrm{x}, \mathrm{x}_n)$

Locally weighted regression

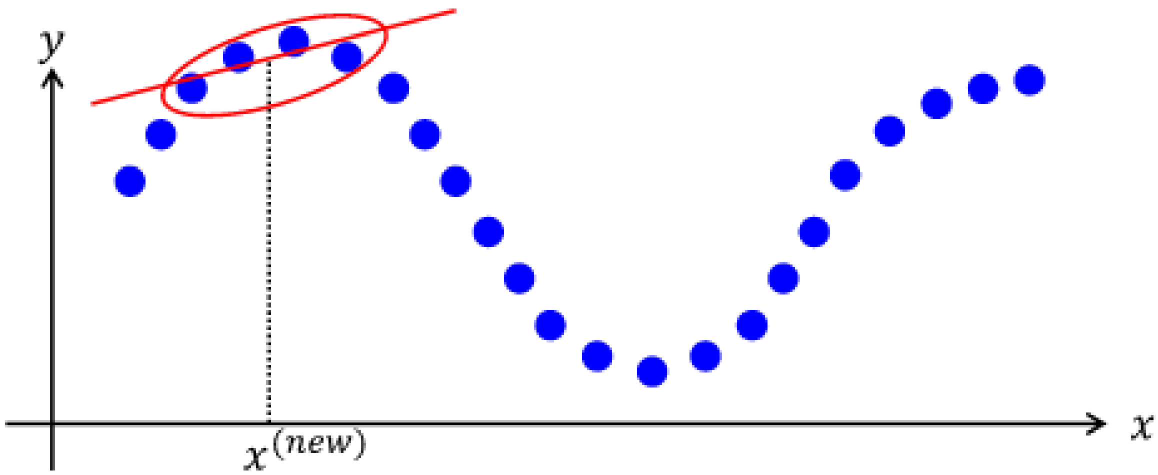
Regression problem $f : X \mapsto \Re$ with data set $D = \{(x_n, t_n)_{n=1}^N\}$

Fit a local regression model around the query sample x_q

1. Compute $N_K(\mathrm{x}_q, D)$: K-nearest neighbors of x_q
2. Fit a regression model $y(\mathrm{x}; \mathrm{w})$ on $N_K(\mathrm{x}_q, D)$
3. Return $y(\mathrm{x}_q; \mathrm{w})$

Locally weighted regression

Example with linear kernel



Summary

1. Non-parametric models based on storing data (lazy approaches)
2. No explicit model
3. Sensitive to parameters and distance function
4. Require storage of all data