Sapienza University of Rome

Master in Artificial Intelligence and Robotics
Master in Engineering in Computer Science

Machine Learning

A.Y. 2019/2020

Prof. L. Iocchi, F. Patrizi, V. Ntouskos

Sapienza University of Rome, Italy - Machine Learning (2019/2020)

10. Instance based learning

L. Iocchi, F. Patrizi, V. Ntouskos

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

K-nearest neighbors

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

Classification with K-NN (target $f : X \mapsto C$, data set $D = \{(x_i, t_i)_{i=1}^n\}$:

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

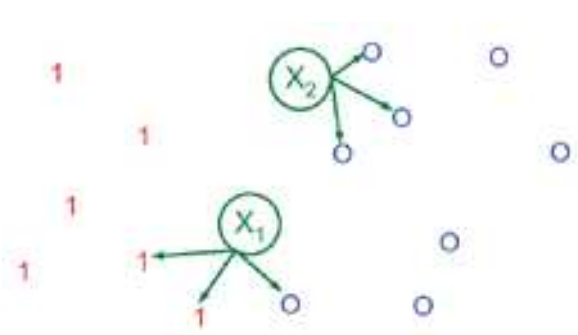
Likelihood of class c for new instance **x**:

$$p(c|\mathbf{x}, D, K) = \frac{1}{K} \sum_{i \in N_K(\mathbf{x}, D)} \mathbb{I}(t_i = c),$$

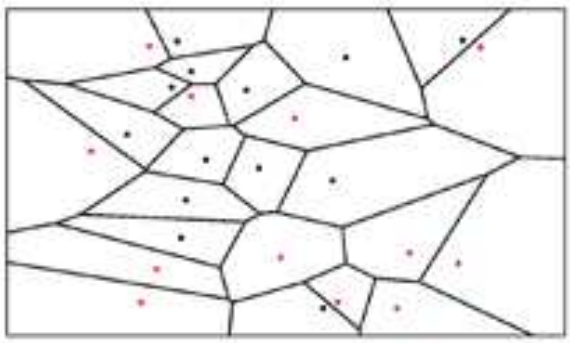
with $N_K(\mathbf{x}, D)$ the K nearest points to **x** and $\mathbb{I}(e) = \begin{cases} 1 & \text{if } e \text{ is true} \\ 0 & \text{if } e \text{ is false} \end{cases}$.

Requires storage of all the data set!

K-nearest neighbors examples



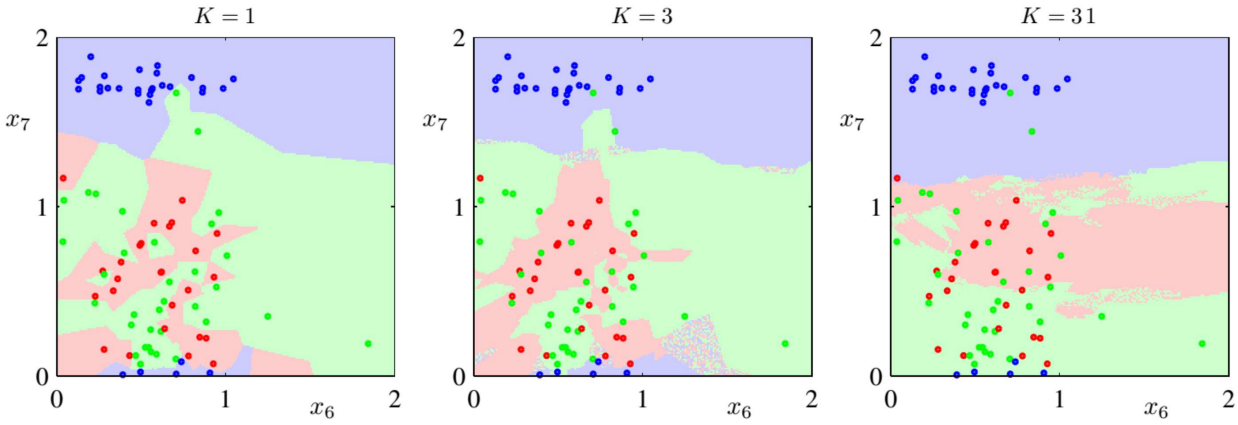
$K = 3$



Voronoi tesselation for $K = 1$

K-nearest neighbors

Increasing K brings to smoother regions (reducing overfitting)



Kernelized nearest neighbors

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

$$\|\mathbf{x} - \mathbf{x}_i\|^2 = \mathbf{x}^T\mathbf{x} + \mathbf{x}_i^T\mathbf{x}_i - 2\mathbf{x}^T\mathbf{x}_i.$$

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

Locally weighted regression

Regression problem $f : X \mapsto \Re$ with data set $D = \{(x_i, t_i)_{i=1}^N\}$

Fit a local regression model around the query sample $\mathbf{x}_q$

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

Summary

1. Non-parametric models based on storing data (lazy approaches)
2. No explicit model
3. Require storage of all data