Sapienza University of Rome

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

Machine Learning

A.Y. 2020/2021

Prof. L. Iocchi, F. Patrizi

10. Instance based learning

L. Iocchi, F. Patrizi

Summary

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

References

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

Sapienza University of Rome, Italy - Machine Learning (2020/2021)

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 $D = \{(\mathbf{x}_n, t_n)_{n=1}^N\}$

Stream 20:19

Classification with K-NN,

1. Find K nearest neighbors of new instance $\mathbf{x}$
2. Assign to $\mathbf{x}$ the most common label among the majority of neighbors

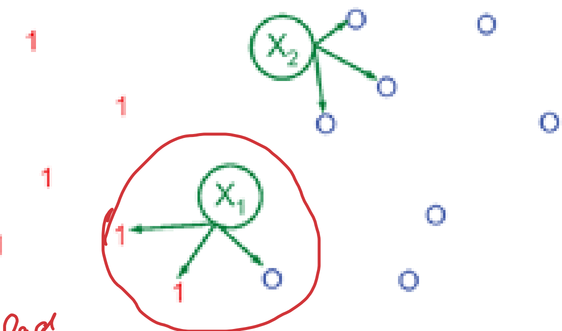
Likelihood of class c for new instance $\mathbf{x}$:

Probability correspond to the density

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

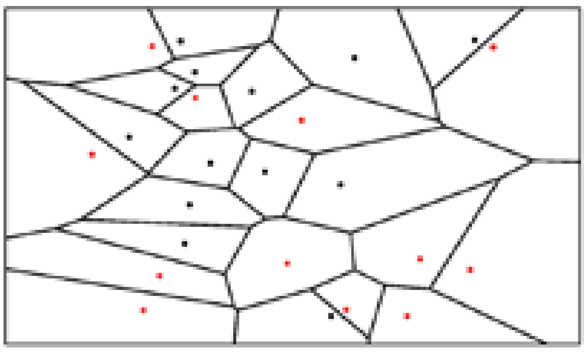
with $N_K(\mathbf{x}_n, D)$ the K nearest points to $\mathbf{x}_n$ and $\mathbb{I}(e) = \begin{cases} 1 & \text{if } e \text{ is true} \\ 0 & \text{if } e \text{ is false} \end{cases}$.

K-nearest neighbors examples



2/3 red
1/3 blue

$K = 3$



Voronoi tesselation for $K = 1$

Requires storage of all the data set!

Depends on a distance function!

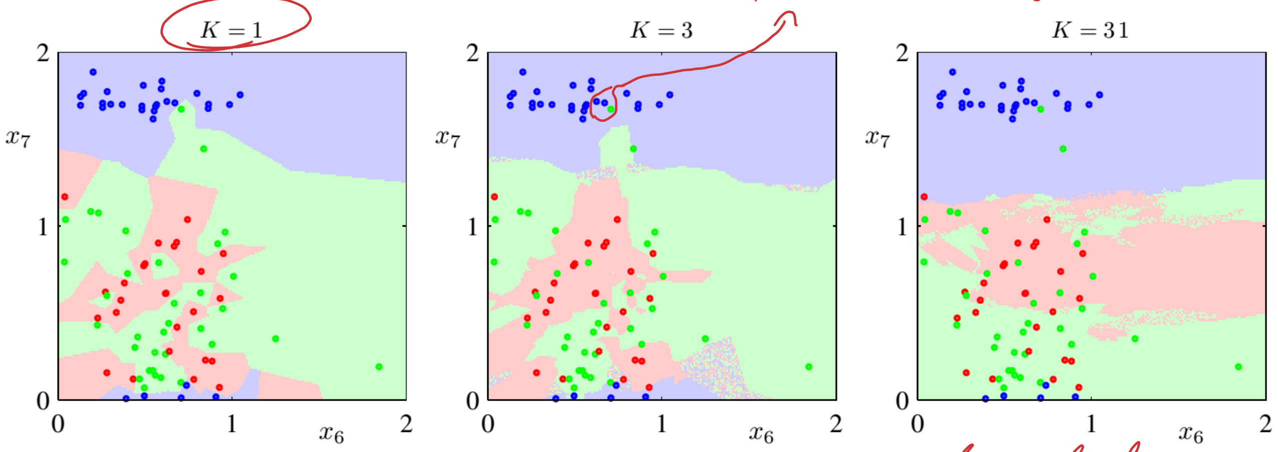
K-nearest neighbors

Don't ensure better performances

Increasing K brings to smoother regions (reducing overfitting)

A point is the closest to itself

Green point is closer to blue than to green



Increase # of misclassified instances.

Kernelized nearest neighbors

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

to find distance from an instance to a point.

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

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

To find solution in a different space

Locally weighted regression

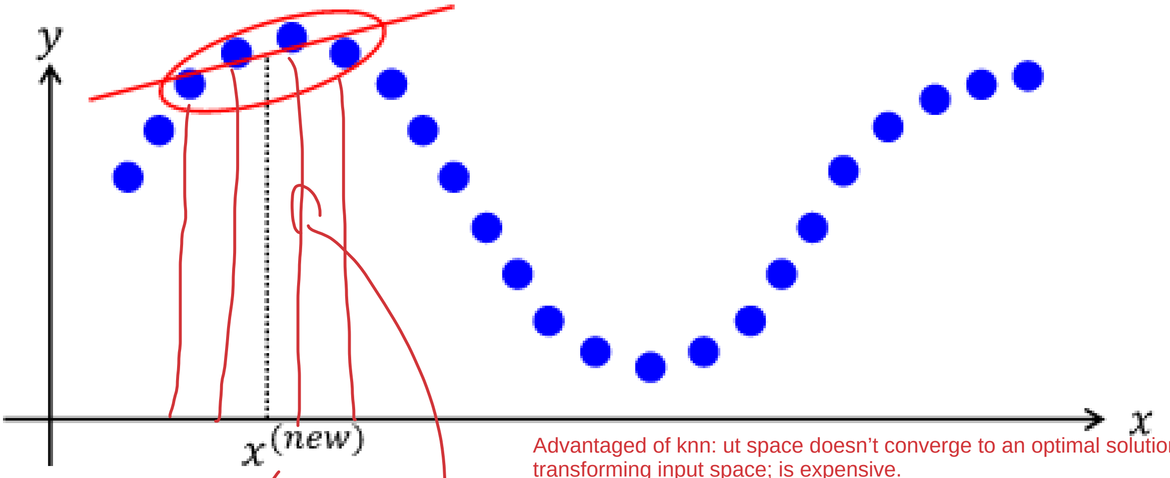
Regression problem $f : X \mapsto \Re$ with data set $D = \{(x_n, t_n)_{n=1}^{N}\}$

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

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 $y(\mathbf{x}_q; \mathbf{w})$

Locally weighted regression

Example with linear kernel



Advantaged of knn: ut space doesn't converge to an optimal solution, so use knn transforming input space; is expensive.

not in Dataset

Compute the nearest

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