Lecture Notes I-2 – Examples of Predictors. Nearest Neighbor and Kernel Predictors. Bias and Variance

Marina Meilă
mmp@uwaterloo.ca

With Thanks to Pascal Poupart & Gautam Kamath
Cheriton School of Computer Science
University of Waterloo

January 12, 2026

Reading HTF Ch.: 2.1–5,2.9, 7.1–4 bias-variance tradeoff, Murphy Ch.: 1., 8.6[1], Bach Ch.:

[1] Neither textbook is close to these notes except in a few places; take them as alternative perspectives or related reading

Training and testing error

- Let $\mathcal{D} = \{(x^1, y^1),\ (x^2, y^2),\ \ldots\ (x^n, y^n)\}$ be the training set and let the K-NN classifier from $\mathcal{D}$ be f_K
- How "good" is f_K?
- **Training error** $= \frac{1}{n}\#(\text{errors of } f_K \text{ on } \mathcal{D}) \;=\; \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[f_K(x^i) \neq y^i]}$

Training and testing error

- Let $\mathcal{D} = \{(x^1, y^1),\ (x^2, y^2),\ \ldots\ (x^n, y^n)\}$ be the training set and let the K-NN classifier from $\mathcal{D}$ be f_K
- How "good" is f_K?
- **Training error** $= \frac{1}{n}\#(\text{errors of } f_K \text{ on } \mathcal{D}) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{[f_K(x^i)\neq y^i]}$
- **Test error** $Pr[f_K(x) \neq y]$

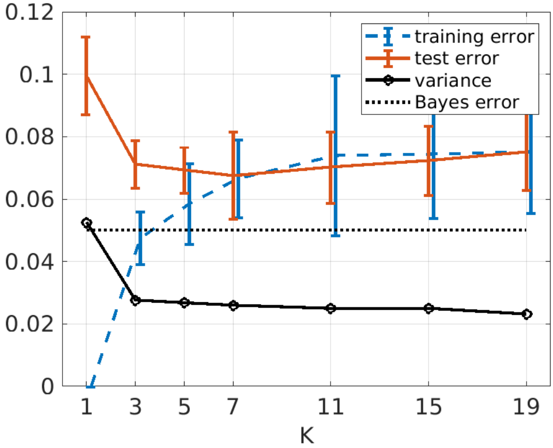
Training and testing error

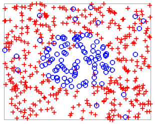
- Let $\mathcal{D} = \{(x^1, y^1), (x^2, y^2), \ldots (x^n, y^n)\}$ be the training set and let the K-NN classifier from $\mathcal{D}$ be f_K
- How "good" is f_K?
- **Training error** $= \frac{1}{n}\#(\text{errors of } f_K \text{ on } \mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[f_K(x^i) \neq y^i]}$
- **Test error** $Pr[f_K(x) \neq y]$ for new points $(x, y) \sim P_{XY}$

Training and testing error

- Let $\mathcal{D} = \{(x^1, y^1),\ (x^2, y^2),\ \ldots\ (x^n, y^n)\}$ be the training set and let the K-NN classifier from $\mathcal{D}$ be f_K
- How "good" is f_K?
- **Training error** $= \frac{1}{n}\#(\text{errors of } f_K \text{ on } \mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[f_K(x^i)\neq y^i]}$
- **Test error** $Pr[f_K(x) \neq y]$ for new points $(x, y) \sim P_{XY}$
- We approximate the test error by using a **test set** $\mathcal{D}^{\text{test}} = \{(\tilde{x}^1, \tilde{y}^1),\ (\tilde{x}^2, \tilde{y}^2), \ldots\ (\tilde{x}^{n'}, \tilde{y}^{n'})\}$ from the same P_{XY}.
- Thus, in practice, **Test error**$= \frac{1}{n}\#(\text{errors of } f_K \text{ on } \mathcal{D}^{\text{test}}) = \frac{1}{n'}\sum_{i=1}^{n'} \mathbf{1}_{[f_K(\tilde{x}^i)\neq \tilde{y}^i]}$

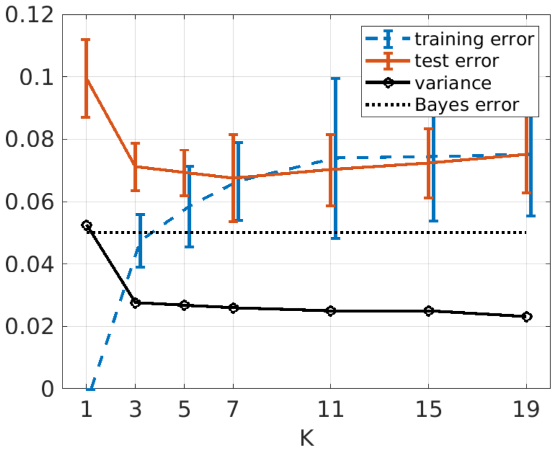
Training and testing error for K-NN

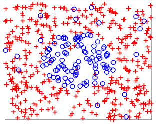




Ignore the "variance" and "Bayes error" for now

Training and testing error for K-NN

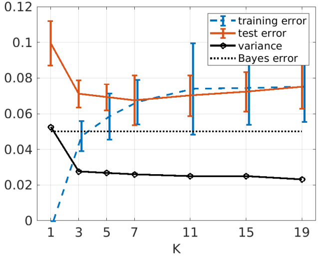




Ignore the "variance" and "Bayes error" for now

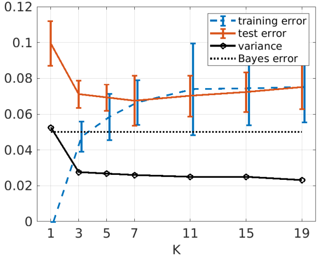
- So, what's happening? For $K = 1$, training error=0 but test error is large
- As K increases, test error decreases at first, then increases again

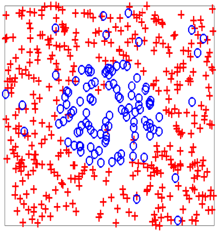
The case $K = 1$: Variance



- $\mathcal{D} \sim P_{XY} \Rightarrow \mathcal{D}$ is **random**

The case $K = 1$: Variance





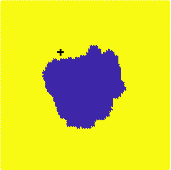
$(K = 1)$

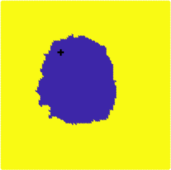
- $\mathcal{D} \sim P_{XY} \Rightarrow \mathcal{D}$ is **random**
- Hence any function f_K we estimate from $\mathcal{D}$ is also random

- Formally, for any fixed x, $f_K(x)$ is a random variable, hence it has a variance.
- In this course, we do not explicitly calculate the variance, but we want to know what increases or decreases it.

The case of K large: Bias

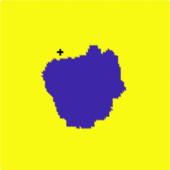
($K = 11$)

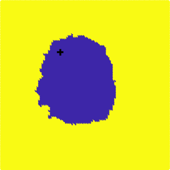




The case of K large: Bias

($K = 11$)





- **Bias** means to let one's own prior beliefs override the evidence.
- In data science/ML/statistics **every model/prediction** is a combination of **prior belief** and data

- **prior** = before seeing the data
- (usually) **prior belief** = prior **knowledge**, e.g. from previous experiments

- Bias can take many forms – in this course you will encounter several
- We do not explicitly calculate bias, but we want to identify where it is coming from, and what increases/decreases it
- One way to look for bias: if a predictor f cannot exactly/accurately predict a training set, "whatever is causing this" is bias.

The Bias-Variance trade-off

- When bias ↗, variance ↘

- When data set size n ↗, variance ↘