STAT406- Methods of Statistical Learning Lecture 14

Matias Salibian-Barrera

UBC - Sep / Dec 2018

© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Classification as prediction

- Most classifiers can be thought of as different ways to estimate or model

$$P\left(G = \mathbf{c}_j \,\middle|\, \mathbf{X} = \mathbf{x}\right)$$

- One way to model these probabilities is via Bayes' Thrm:

$$P\left(G = \mathbf{c}_j \,\middle|\, \mathbf{X} = \mathbf{x}\right) = \mathbf{f}\left(\mathbf{x} \,\middle|\, \mathbf{c}_j\right) P\left(G = \mathbf{c}_j\right) \big/ \mathbf{f}\left(\mathbf{x}\right)$$

© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

LDA vs QDA

LDA classifies an observation into the class c_j for which the estimated (under a normality assumption)

$$P\left(G = c_j \middle| \mathbf{X} = \mathbf{x}\right)$$

is highest.

If $\mathbf{X} | G = c_j \sim \mathcal{N}\left(\mu_j, \mathbf{\Sigma}\right)$ it means the class for which

$$\delta_j(\mathbf{x}) = \mathbf{x}' \, \mathbf{a}_j + b_j$$

is highest, where

$$\mathbf{a}_j = \mathbf{\Sigma}^{-1} \mu_j \quad \text{and} \quad b_j = -\frac{1}{2} \mu_j' \mathbf{\Sigma}^{-1} \mu_j + \log(p_j)$$

© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

QDA

If $\mathbf{X} | G = c_j \sim \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ then, the estimated posterior probabilities look different.

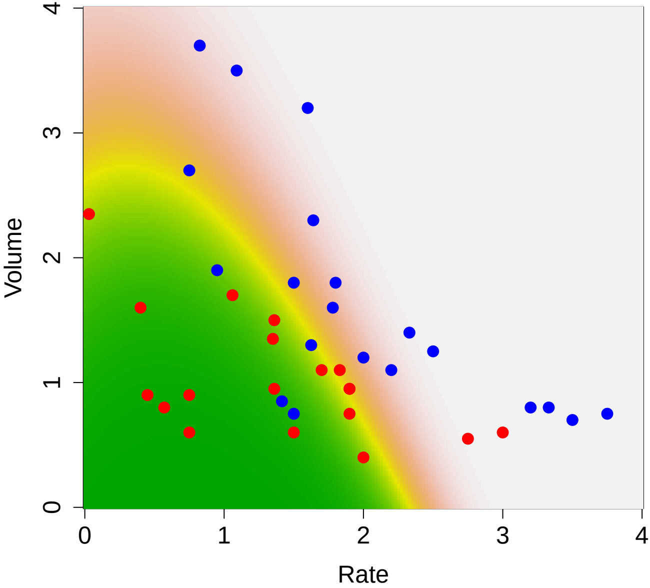
QDA classifies an observation into the class c_j that maximizes

$$\delta_j(\mathbf{x}) = \mathbf{x}' \mathbf{A}_j \mathbf{x} + \mathbf{x}' \mathbf{d}_j + b_j$$

(see the book for details)

© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

QDA-based probabilities



© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

LDA - QDA - Logistic classifiers

- Multiclass: LDA, QDA and Logistic classifiers extend to the case of more than 2 classes

- Examples in R (Zip-code hand-written digits)

© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

LDA - QDA - Logistic classifiers

- LDA vs QDA: presents the usual "flexibility vs. variability" trade-off

- LDA & QDA vs Logistic classifiers: **Gaussian MLE estimates** (non-robust, sensitive to the Gaussian assumption) vs. **Binomial MLE estimates** (no distributional assumption of $\mathbf{X} | G = c_j$ required).

© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Nearest-neighbours

- We need to estimate

$$P\left(G = \mathbf{g} \middle| \mathbf{X} = \mathbf{x}\right)$$

- An intuitive and “model-free” estimator is the nearest-neighbours estimator

© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Nearest-neighbours

- Same spirit as the local-constant (kernel) regression estimator
- The K-NN estimator is

$$\hat{P}\left(G = \mathbf{g} \,|\, \mathbf{X} = \mathbf{x}\right) = \frac{1}{|N_{\mathbf{x}}^{K}|} \sum_{j \in N_{\mathbf{x}}^{K}} \mathbf{I}\left\{Y_j = \mathbf{g}\right\}$$

where

$$N_{\mathbf{x}}^{K} = \left\{i \, : \, d(\mathbf{X}_i, \mathbf{x}) \leq d_{(K)}\right\}$$

and $d_{(K)}$ is the distance from $\mathbf{x}$ to the K-th closest point in the sample ($\mathbf{X}_i$)

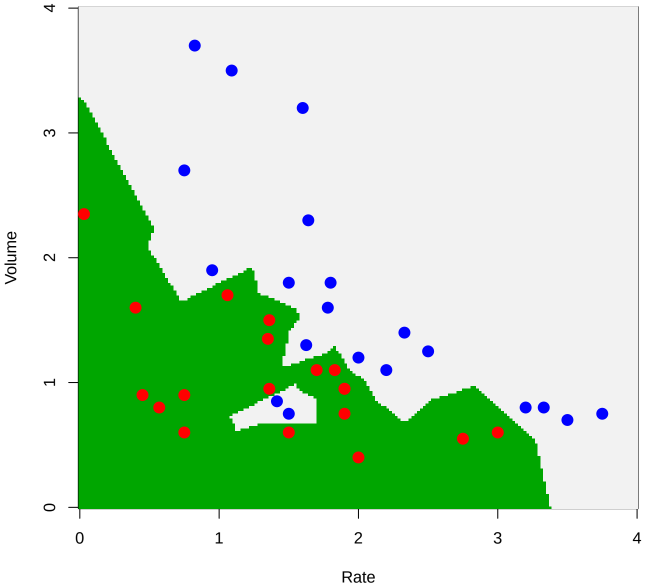
© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Nearest-neighbours

- The K-NN estimator is the proportion of observations from class **g** among the closest K neighbours
- The K-NN classifier assigns a point to the class most represented among its K neighbours (“peer pressure”)

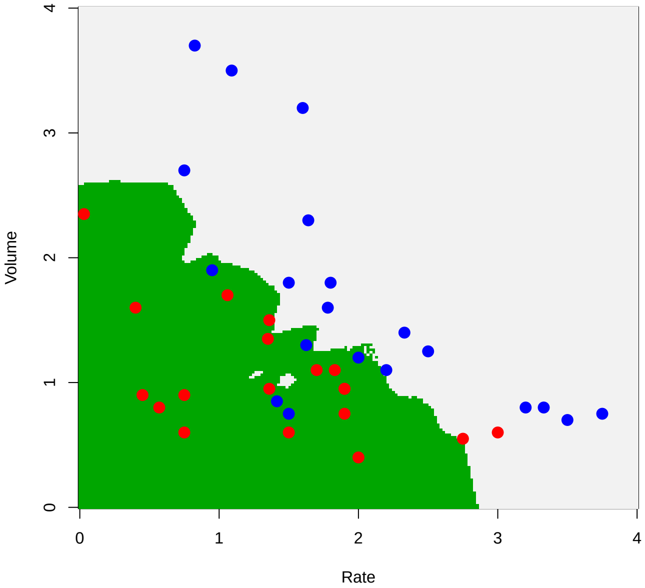
© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Vaso – 1-NN



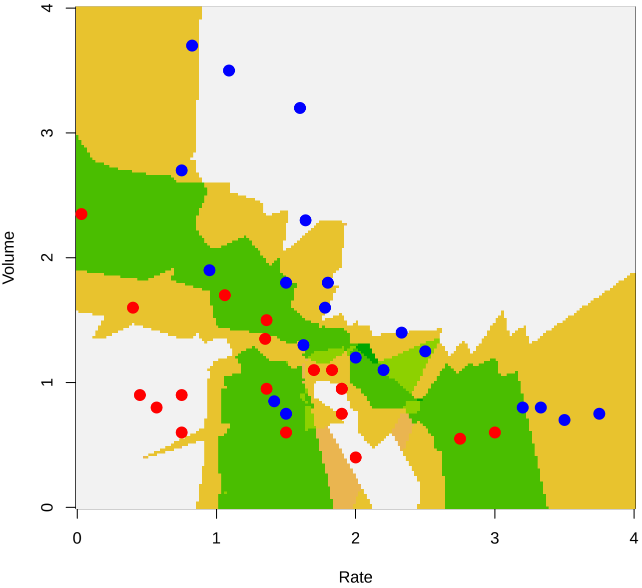
© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Vaso – 5-NN



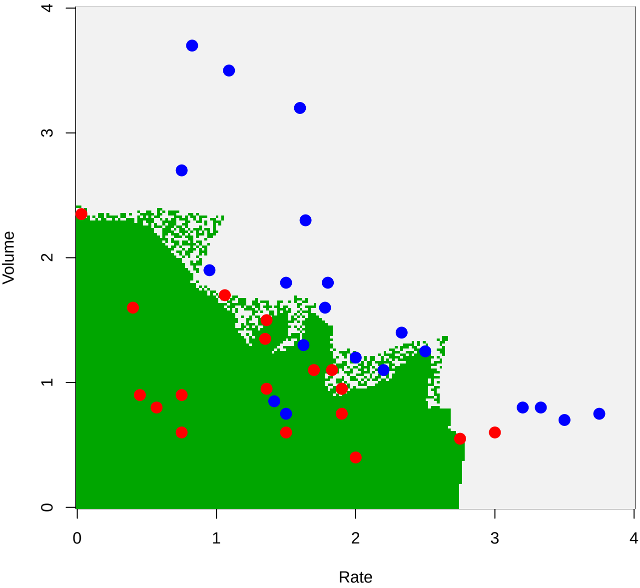
© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Vaso – 5-NN - votes



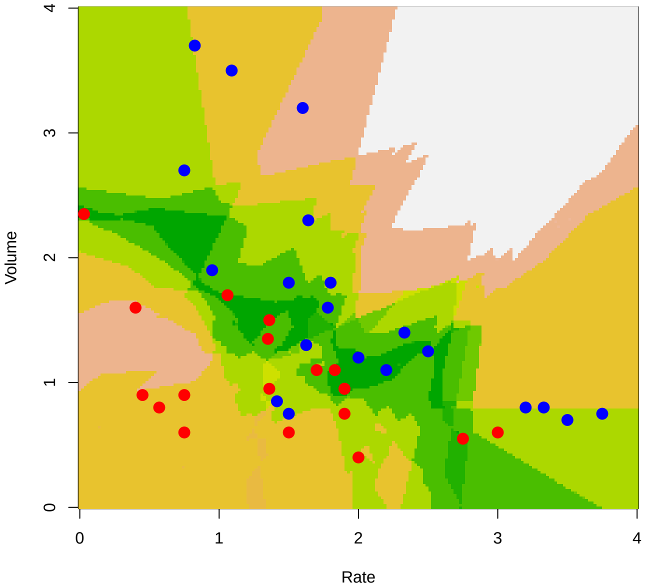
© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Vaso – 10-NN



© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Vaso – 10-NN - votes



© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

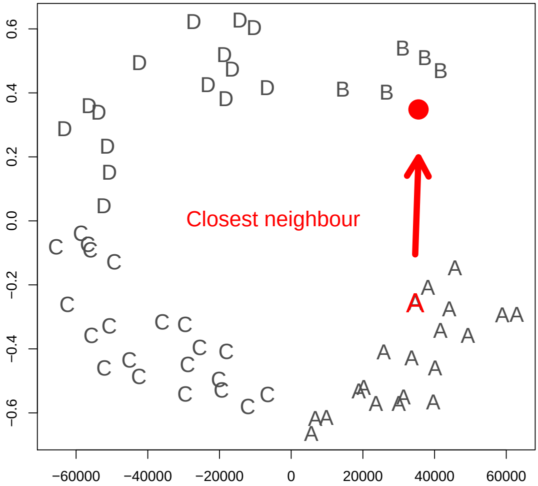
Nearest-neighbours

- How can we select the number K of neighbours?
- Zip-code example

© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.

Nearest-neighbours - Challenge

- What’s wrong with this picture?



© Matias Salibian-Barrera, 2018. All rights reserved. Cannot be copied, re-used, or edited.