Classification

Bernhard Stankewitz
bernhard.stankewitz@posteo.de

February 18, 2019

Abstract

...

Contents

1 Elementary definitions and the Bayes classifier

Definition 1.1 (Classifier)**.**

(a) For i.i.d. training data $(X_1, Y_1), \dots, (X_n, Y_n) \in \mathbb{R}^d \times \{1, \dots, K\}$, a classifier is a measurable function $C : \mathbb{R}^d \to \{1, \dots, K\}$. The classification error is given by

$$R(C) := \mathbb{P}\{C(X) \neq Y\} = \mathbb{E}\mathbf{1}_{C(X) \neq Y}.$$

(b) In case that the labels are given by $\{0, 1\}$, the classification error

$$R(C) = \mathbb{E}(Y - C(X))^2 = \int \mathbf{1}_{y \neq C(x)} \, \mathbb{P}^{X,Y}(d(x, y)).$$

In case all theoretical quantities are known, a classification problem has an optimal solution.

Proposition 1.2 (Bayes-Classifier)**.**

(i) In the situation of Definition 1.1, the classification error is minimised by the Bayes classifier

$$C^{Bayes}(x) := \operatorname*{argmax}_{k=1,\dots,K} \mathbb{P}\{Y = k | X = x\}.$$

(ii) If the labels are given by $\{0, 1\}$, we have

$$C^{Bayes}(x) = \mathbf{1}\{\eta(x) \geq 1/2\} \qquad \text{with} \qquad \eta(x) := \mathbb{P}\{Y = 1 | X = x\}.$$

Proof. For any classifier C, we have

$$R(C) = 1 - \mathbb{E}\mathbb{E}\mathbf{1}_{C=Y}|X = 1 - \mathbb{E}\sum_{k=1}^{K} \mathbb{E}(\mathbf{1}_{C=k}\mathbf{1}_{Y=k}|X) = 1 - \mathbb{E}\sum_{k=1}^{K} \mathbf{1}_{C=k}\mathbb{E}(\mathbf{1}_{Y=k}|X).$$

<++> □

2 The KNN-classifier

Definition 2.1 (KNN-classifier)**.**

(a) Let $(X_1, Y_1), \dots, (X_n, Y_n) \in \mathbb{R}^d \times \{1, \dots, J\}$ be a training sample and $K \in \mathbb{N}$. For $x \in \mathbb{R}^d$, let $N_K(x)$ be the set of the K nearest neighbours of x with respect to the euclidean distance. Then, the KNN-classifier is given by

$$\hat{C}^{\mathrm{KNN}}(x) := \operatorname*{argmax}_{j=1,\dots,J} \frac{1}{K} \sum_{X_i \in N_K(x)} \mathbf{1}_{Y_i = j}.$$

(b) In case, the labels are given by $\{0, 1\}$, we have

$$\hat{C}^{\mathrm{KNN}}(x) = \mathbf{1}\{\hat{\eta}(x) \geq 1/2\} \qquad \text{with} \qquad \hat{\eta}(x) := \frac{1}{K} \sum_{X_i \in N_K(x)} \mathbf{1}_{Y_i = 1} =: \sum_{i=1}^{n} w_i(x) Y_i,$$

where $w_i := \mathbf{1}_{X_i \in N_K(x)}/K$ with $\sum_{i=1}^n w_i = 1$.

Theorem 2.2 (Consistency of KNN)**.**