Universität Salzburg | Florian Graf

Machine Learning

Übungsblatt **3** | 23.5 Punkte

Aufgabe 1. 11 P.

Es seien reellwertige Beobachtungen $x_1, \ldots, x_n \in \mathbb{R}$ mit Klassenlabel $y_1, \ldots, y_n \in \{0, 1\}$ gegeben. Wir möchten auf Basis der Beobachtungen ein Entscheidungsmodell erstellen. Dazu nehmen wir an, dass $p(y = c|\lambda) = \mathrm{Ber}(y|\lambda)$ Bernoulli verteilt und dass $p(\mathbf{x}|y = 0) = \mathcal{N}(\mathbf{x}|\mu_0, \sigma_0^2)$ und $p(\mathbf{x}|y = 1) = \mathcal{N}(\mathbf{x}|\mu_1, \sigma_1^2)$ univariat normalverteilt sind. Das Ziel ist es die unbekannten Verteilungensparameter $\lambda, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2$ zu bestimmen, bzw. zu schätzen.

(a) Geben Sie die gemeinsame Verteilung $p(y_i, x_i|\lambda, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2)$ von x_i und y_i an. Nutzen Sie diese um die gemeinsame Verteilung $p(\mathcal{D}|\lambda, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2)$ von $\mathcal{D} = \{(y_1, x_1), \ldots, (y_n, x_n)\}$ zu bestimmen, unter der Annahme, dass $(x_1, y_1), \ldots, (x_n, y_n)$ gemeinsam unabhängig sind.

Die Funktion

$$L_{\mathcal{D}} : [0, 1] \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^+$$
$$(\lambda, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2) \mapsto p(\mathcal{D}|\lambda, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2)$$

wird Likelihoodfunktion genannt. Bestimmen Sie ein Tupel an Parametern $(\hat{\lambda}, \hat{\mu}_0, \hat{\sigma}_0^2, \hat{\mu}_1, \hat{\sigma}_1^2) = \arg\max L_{\mathcal{D}}$ sodass die Likelihoodfunktion maximal ist.

(b) Zeigen Sie zunächst, dass die Parameter $\hat{\lambda}$ und $(\hat{\mu}_0, \hat{\sigma}_0^2)$ und $(\hat{\mu}_1, \hat{\sigma}_1^2)$ separat bestimmt werden können.

(c) Zeigen Sie, dass

$$\hat{\lambda} = \frac{n_1}{n}, \quad n_1 = |\{i : y_i = 1\} \ , \tag{1}$$

indem Sie ein lokales Maximum des zugehörigen Summanden der log Likelihoodfunktion $\log L_{\mathcal{D}}$ bestimmen.

(d) Zeigen Sie, dass (für $c = 0, 1$),

$$\hat{\mu}_c = \frac{1}{n_c} \sum_{i:y_i=c} x_i, \quad \hat{\sigma}_i^2 = \frac{1}{n_c} \sum_{i:y_i=c} (x_i - \hat{\mu}_c)^2 \ . \tag{2}$$

indem Sie ein lokales Maximum des zugehörigen Summanden der log Likelihoodfunktion $\log L_{\mathcal{D}}$ bestimmen.

Aufgabe 2. 5 P.

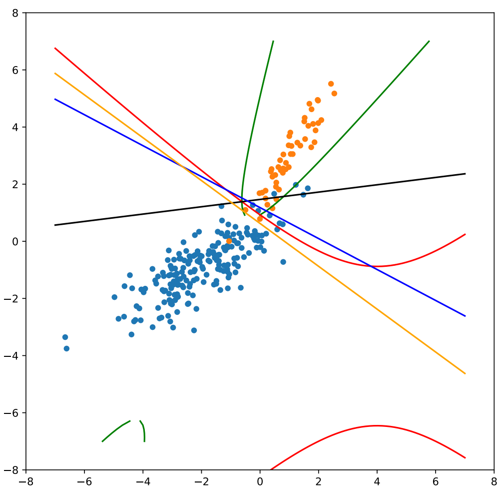
In dem Setting von Aufgabe 1 ändern wir nun die Verteilungsannahmen.

(a) Die a priori Klassenverteilung ist uniform, d.h. $p(y = 0) = p(y = 1)$. Begründen Sie, welche Verteilungsparameter gleich wie in Aufgabe 1 geschätzt werden können. Leiten Sie für die übrigen Parameter die Maximum-Likelihood Schätzer her.

(b) Die bedingten Verteilungen $p(x|y = 0, \mu_0, \sigma_0^2)$ und $p(x|y = 1, \mu_1, \sigma_1^2)$ erfüllen die 'tied-covariances' Annahme. Begründen Sie, welche Verteilungsparameter gleich wie in Aufgabe 1 geschätzt werden können. Leiten Sie für die übrigen Parameter die Maximum-Likelihood Schätzer her.

Aufgabe 3. 7.5 P.

Die nachfolgende Abbildung zeigt eine Stichprobe zweier Klassen und Entscheidungsgrenzen, die durch verschiedene Varianten der Gaußschen Diskriminanzanalysis (mit Maximum-Likelihood Schätzung) bestimmt wurden.



Ordnen Sie die Entscheidungsgrenzen den folgenden Klassifizierungsmodellen zu. Begründen Sie Ihre Entscheidung und nennen Sie das zugrunde liegende Wahrscheinlichkeitsmodell.

(i) GDA (allg. Fall / ohne Einschränkungen)

(ii) Naive Bayes (mit Gaußverteilung)

(iii) LDA

(iv) Diagonal LDA

(v) Nearest Centroid