Gaußsche Diskriminanzanalyse

Dr. rer. nat Dennis Müller

January 18, 2016

Table of Contents

1 Quadratische Diskriminanzanalyse

2 Gaußsche Diskriminanzanalyse

Einführung

Quadratische Diskriminanzanalyse - Merkmalsklassifikation

QDA

- Wir betrachten zunächst ein *Zwei-Klassen* Klassifikationsproblem.
- Dabei sei zu einem zu klassifizierenden Objekt ein n-dimensionaler *Featurevektor* $\mathbf{x}_i \in \mathbb{R}^n$ gegeben. Wir nehmen an dieser Vektor entstammt einer von zwei Klassen.
- Mit $y_i \in \{0, 1\}$ bezeichnen wir die Klasse des i-ten Featurevektors.
- Nach dem Satz von Bayes und dem Satz der totalen Wahrscheinlichkeit folgt dann

$$P(y_i|\mathbf{x}_i) = \frac{P(\mathbf{x}_i|y_i) \cdot p(y_i)}{P(\mathbf{x}_i|y_i = 0) \cdot p(y_i = 0) + P(\mathbf{x}_i|y_i = 1) \cdot p(y_i = 1)}$$

QDA

- In der Quadratische Diskriminanzanalyse gehen wir davon aus, dass die Featurevektoren innerhalb einer Klasse einer Normalverteilung entsprechen, also

$$P(\mathbf{x}_i|y_i) = \frac{1}{\sqrt{(2\pi)^n|\Sigma_{y_i}|}} \exp\left(-\frac{1}{2}\left(\mathbf{x}_i - \mu_{y_i}\right)^T \Sigma_{y_i}^{-1}\left(\mathbf{x}_i - \mu_{y_i}\right)\right)$$

- Ferner nehmen wir an das die *a-priori* Wahrscheinlichkeit für die Klassenzugehörigkeit Bernoulli-verteilt ist, also

$$P(y_i) = \theta^{y_i} \cdot (1-\theta)^{(1-y_i)}$$

  mit $\theta \in [0,1]$.
- Die Parameter unseres Modells sind also $\theta, \mu_0, \mu_1, \Sigma_0$ und Σ_1.

QDA

- Angenommen wir haben ein Trainingsset von k Featurevektoren mit bekannter Klassenzugehörigkeit. Die *log-likelihood* der Daten ist dann

$$
\begin{aligned}
\lambda(\theta, \mu_0, \mu_1, \Sigma_0, \Sigma_1) &= \sum_{i=1}^{k} \log\ P(x_i, y_i | \theta, \mu_0, \mu_1, \Sigma_0, \Sigma_1) \\
&= \sum_{i=1}^{k} \log\ P(x_i | y_i, \mu_0, \mu_1, \Sigma_0, \Sigma_1) \cdot P(y_i | \theta)
\end{aligned}
$$

QDA

- Die *Maximum-likelihood* Schätzung für die fünf Parameter ergibt sich dann zu

$$\theta = \frac{1}{k}\sum_{i=1}^{k} y_i$$

$$\mu_0 = \frac{\sum_{i=1}^{k} x_i(1-y_i)}{\sum_{i=1}^{k}(1-y_i)}$$

$$\mu_1 = \frac{\sum_{i=1}^{k} x_i y_i}{\sum_{i=1}^{k} y_i}$$

$$\Sigma_0 = \frac{\sum_{i=1}^{k}(x_i-\mu_0)(x_i-\mu_0)^T(1-y_i)}{\sum_{i=1}^{k}(1-y_i)}$$

$$\Sigma_1 = \frac{\sum_{i=1}^{k}(x_i-\mu_1)(x_i-\mu_1)^T y_i}{\sum_{i=1}^{k} y_i}$$

QDA

- Sind die Parameter nun bekannt, z.B. durch obige ML-Schätzung, können wir einen neuen Featurevektor mit unbekannter Klassenzugehörigkeit *klassifizieren*. Dazu betrachten wir zunächst das Verhältnis der *a-posteriori* Wahrscheinlichkeiten

$$\frac{P(y_i = 1|\mathbf{x}_i)}{P(y_i = 0|\mathbf{x}_i)} = \frac{P(\mathbf{x}_i|y_i = 1)P(y_i = 1)}{P(\mathbf{x}_i|y_i = 0)P(y_i = 0)}$$

- Auch hier betrachten wir wieder den Logarithmus und erhalten zunächst

$$\log P(\mathbf{x}_i|y_i) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_{y_i}| - \frac{1}{2}(\mathbf{x}_i - \mu_{y_i})^T \Sigma_{y_i}^{-1}(\mathbf{x}_i - \mu_{y_i})$$

und

$$\log P(y_i) = y_i \log\theta + (1 - y_i)\log(1 - \theta)$$

QDA

- Nun gilt für den Logarithmus des Verhältnisses

$$g(\mathbf{x}_i) = \log \frac{P(y_i=1|\mathbf{x}_i)}{P(y_i=0|\mathbf{x}_i)} = -\frac{1}{2}\log|\Sigma_1| + \frac{1}{2}\log|\Sigma_0| -\frac{1}{2}(\mathbf{x}_i-\mu_0)^T\Sigma_0^{-1}(\mathbf{x}_i-\mu_0) + \frac{1}{2}(\mathbf{x}_i-\mu_1)^T\Sigma_1^{-1}(\mathbf{x}_i-\mu_1) + log\left(\frac{\theta}{1-\theta}\right)$$

- g heisst *Entscheidungsfunktion*.

QDA

- Der mittlere Term lässt sich schreiben als (aus-multiplizieren und neu zusammenfassen)

$$\frac{1}{2}\mathbf{x}_i^T\left(\Sigma_1^{-1}-\Sigma_0^{-1}\right)\mathbf{x}_i$$
$$+$$
$$\mathbf{x}_i^T\left(\Sigma_0^{-1}\mu_0-\Sigma_1^{-1}\mu_1\right)$$
$$+$$
$$\frac{1}{2}\mu_1^T\Sigma_1^{-1}\mu_1-\frac{1}{2}\mu_0^T\Sigma_0^{-1}\mu_0$$

Gaußsche Diskriminanzanalyse - Spezialfall

GDA

- Bei der quadratischen Diskriminanzanalyse (QDA) gehen wir davon aus, das beide Klassen unterschiedliche Varianzen haben können. Dadurch ergibt sich eine, i.d.r. quadratische *Grenzfläche* zwischen den beide Klassen durch den Term

$$\mathbf{x}_i^T \left(\Sigma_1^{-1} - \Sigma_0^{-1}\right) \mathbf{x}_i$$

.

GDA

- Bei der Gaußschen Diskriminanzanalyse (GDA) nehmen wir an, das beide Klassen *die gleiche* Varianz haben, also $\Sigma_0 = \Sigma_1 = \Sigma$. Dann ergibt sich für die ML Schätzung

$$\Sigma = \frac{1}{k}\sum_{i=1}^{k}\left(\mathbf{x}_i - \mu_{y_i}\right)\left(\mathbf{x}_i - \mu_{y_i}\right)^T$$

- Für die *Entscheidungsfunktion* folgt dann

$$g(\mathbf{x}_i) = x_i^T\Sigma^{-1}(\mu_0-\mu_1)+\frac{1}{2}(\mu_1-\mu_0)^T\Sigma^{-1}(\mu_1-\mu_0)+\log\left(\frac{\theta}{1-\theta}\right)$$

- Hier ist die Entscheidungsfunktion also *linear* in $\mathbf{x}_i$.

GDA

- Für $\theta = 1/2$, also a-priori gleich wahrscheinliche Klassen ist $\log(\theta/(1-\theta)) = 0$. Die Entscheidungsfunktion betrachtet dann nur die mit der (gleichen) Kovarianz gewichtete Distanz zwischen dem Featurevektor und beiden Klassenzentren. Die Zuordnung zu Klasse 1 erfolgt dann also g.d.w.

$$(\mathbf{x}_i - \mu_1)^T \Sigma^{-1} (\mathbf{x}_i - \mu_1) \leq (\mathbf{x}_i - \mu_0)^T \Sigma^{-1} (\mathbf{x}_i - \mu_0)$$

- Die so gewichtete Distanz bezeichnet man auch als *Mahalanobis*-Distanz.
- In diesem Spezialfall ist die GDA also ein Nearest-Neighbor Klassifikator.