These are the slides of the lecture

Pattern Recognition
Winter term 2020/21
Friedrich-Alexander University of Erlangen-Nuremberg.

Erlangen, January 8, 2021
Prof. Dr.-Ing. Andreas Maier

Pattern Recognition (PR)

Prof. Dr.-Ing. Andreas Maier
Pattern Recognition Lab (CS 5), Friedrich-Alexander-Universität Erlangen-Nürnberg
Winter Term 2020/21

Logistic Regression I

Logistic Regression

Logistic Regression is a discriminative model, because it models the posterior probabilities $p(y|\boldsymbol{x})$ directly.

Posteriors and the Logistic Function

For two classes $y \in \{0, 1\}$ we get:

$$p(y = 0|\boldsymbol{x}) \quad =$$

Posteriors and the Logistic Function

For two classes $y \in \{0, 1\}$ we get:

$$p(y = 0|\boldsymbol{x}) \quad = \quad \frac{p(y = 0) \cdot p(\boldsymbol{x}|y = 0)}{p(\boldsymbol{x})}$$

Posteriors and the Logistic Function

For two classes $y \in \{0, 1\}$ we get:

$$
\begin{aligned}
p(y=0|\boldsymbol{x}) &= \frac{p(y=0)\cdot p(\boldsymbol{x}|y=0)}{p(\boldsymbol{x})} \\
&= \frac{p(y=0)\cdot p(\boldsymbol{x}|y=0)}{p(y=0)p(\boldsymbol{x}|y=0)+p(y=1)p(\boldsymbol{x}|y=1)}
\end{aligned}
$$

Posteriors and the Logistic Function

For two classes $y \in \{0, 1\}$ we get:

$$
\begin{aligned}
p(y=0|\boldsymbol{x}) &= \frac{p(y=0)\cdot p(\boldsymbol{x}|y=0)}{p(\boldsymbol{x})} \\
&= \frac{p(y=0)\cdot p(\boldsymbol{x}|y=0)}{p(y=0)p(\boldsymbol{x}|y=0)+p(y=1)p(\boldsymbol{x}|y=1)} \\
&= \frac{1}{1+\frac{p(y=1)p(\boldsymbol{x}|y=1)}{p(y=0)p(\boldsymbol{x}|y=0)}}
\end{aligned}
$$

Posteriors and the Logistic Function (cont.)

$$p(y = 0|\boldsymbol{x}) = \frac{1}{1 + \frac{p(y=1)p(\boldsymbol{x}|y=1)}{p(y=0)p(\boldsymbol{x}|y=0)}}$$

Posteriors and the Logistic Function (cont.)

$$p(y=0|\boldsymbol{x}) \quad = \quad \frac{1}{1+\frac{p(y=1)p(\boldsymbol{x}|y=1)}{p(y=0)p(\boldsymbol{x}|y=0)}}$$

(Trick: extend with exponential and logarithm)

Posteriors and the Logistic Function (cont.)

$$p(y=0|\boldsymbol{x}) = \frac{1}{1+\frac{p(y=1)p(\boldsymbol{x}|y=1)}{p(y=0)p(\boldsymbol{x}|y=0)}}$$

(Trick: extend with exponential and logarithm)

$$= \frac{1}{1+e^{\log\frac{p(y=1)p(\boldsymbol{x}|y=1)}{p(y=0)p(\boldsymbol{x}|y=0)}}}$$

Posteriors and the Logistic Function (cont.)

$$p(y=0|\boldsymbol{x}) = \frac{1}{1+\frac{p(y=1)p(\boldsymbol{x}|y=1)}{p(y=0)p(\boldsymbol{x}|y=0)}}$$

(Trick: extend with exponential and logarithm)

$$= \frac{1}{1+e^{\log\frac{p(y=1)p(\boldsymbol{x}|y=1)}{p(y=0)p(\boldsymbol{x}|y=0)}}}$$

$$= \frac{1}{1+e^{-\log\frac{p(y=0)}{p(y=1)}-\log\frac{p(\boldsymbol{x}|y=0)}{p(\boldsymbol{x}|y=1)}}}$$

Posteriors and the Logistic Function (cont.)

$$p(y=0|\boldsymbol{x}) = \frac{1}{1+\frac{p(y=1)p(\boldsymbol{x}|y=1)}{p(y=0)p(\boldsymbol{x}|y=0)}}$$

(Trick: extend with exponential and logarithm)

$$= \frac{1}{1+e^{\log\frac{p(y=1)p(\boldsymbol{x}|y=1)}{p(y=0)p(\boldsymbol{x}|y=0)}}}$$

$$= \frac{1}{1+e^{-\log\frac{p(y=0)}{p(y=1)}-\log\frac{p(\boldsymbol{x}|y=0)}{p(\boldsymbol{x}|y=1)}}}$$

$$= \frac{1}{1+e^{-\log\frac{p(y=0|\boldsymbol{x})}{p(y=1|\boldsymbol{x})}}}$$

Posteriors and the Logistic Function (cont.)

We see that the posterior for class $y = 0$ can be written in terms of a logistic function:

$$p(y = 0|\boldsymbol{x}) = \frac{1}{1 + e^{-F(\boldsymbol{x})}}$$

Posteriors and the Logistic Function (cont.)

We see that the posterior for class $y = 0$ can be written in terms of a logistic function:

$$p(y = 0|\boldsymbol{x}) \quad = \quad \frac{1}{1 + e^{-F(\boldsymbol{x})}}$$

And thus the posterior for the other class $y = 1$:

$$p(y = 1|\boldsymbol{x}) \quad =$$

Posteriors and the Logistic Function (cont.)

We see that the posterior for class $y = 0$ can be written in terms of a logistic function:

$$p(y = 0|\boldsymbol{x}) = \frac{1}{1 + e^{-F(\boldsymbol{x})}}$$

And thus the posterior for the other class $y = 1$:

$$p(y = 1|\boldsymbol{x}) = 1 - p(y = 0|\boldsymbol{x})$$

Posteriors and the Logistic Function (cont.)

We see that the posterior for class $y = 0$ can be written in terms of a logistic function:

$$p(y = 0|\boldsymbol{x}) = \frac{1}{1 + e^{-F(\boldsymbol{x})}}$$

And thus the posterior for the other class $y = 1$:

$$\begin{aligned} p(y = 1|\boldsymbol{x}) &= 1 - p(y = 0|\boldsymbol{x}) \\ &= \frac{e^{-F(\boldsymbol{x})}}{1 + e^{-F(\boldsymbol{x})}} \end{aligned}$$

Posteriors and the Logistic Function (cont.)

We see that the posterior for class $y = 0$ can be written in terms of a logistic function:

$$p(y = 0|\boldsymbol{x}) = \frac{1}{1 + e^{-F(\boldsymbol{x})}}$$

And thus the posterior for the other class $y = 1$:

$$\begin{aligned} p(y = 1|\boldsymbol{x}) &= 1 - p(y = 0|\boldsymbol{x}) \\ &= \frac{e^{-F(\boldsymbol{x})}}{1 + e^{-F(\boldsymbol{x})}} \\ &= \frac{1}{1 + e^{F(\boldsymbol{x})}} \end{aligned}$$

Posteriors and the Logistic Function (cont.)

Definition

The *logistic function* (also called *sigmoid function*) is defined by

$$g(x) = \frac{1}{1+e^{-x}}$$

where $x \in \mathbb{R}$.

Posteriors and the Logistic Function (cont.)

The derivative of the sigmoid function fulfills the nice property:

$$g'(x) = \left(\frac{1}{1+e^{-x}}\right)' =$$

Posteriors and the Logistic Function (cont.)

The derivative of the sigmoid function fulfills the nice property:

$$g'(x) \quad = \quad \left(\frac{1}{1+e^{-x}}\right)' = \left((1+e^{-x})^{-1}\right)' =$$

Posteriors and the Logistic Function (cont.)

The derivative of the sigmoid function fulfills the nice property:

$$g'(x) = \left(\frac{1}{1+e^{-x}}\right)' = \left((1+e^{-x})^{-1}\right)' = \frac{1}{(1+e^{-x})^2} \cdot e^{-x}$$

Posteriors and the Logistic Function (cont.)

The derivative of the sigmoid function fulfills the nice property:

$$
\begin{aligned}
g'(x) &= \left(\frac{1}{1+e^{-x}}\right)' = \left((1+e^{-x})^{-1}\right)' = \frac{1}{(1+e^{-x})^2} \cdot e^{-x} \\
&= \frac{1}{(1+e^{-x})} \cdot \frac{e^{-x}}{(1+e^{-x})}
\end{aligned}
$$

Posteriors and the Logistic Function (cont.)

The derivative of the sigmoid function fulfills the nice property:

$$
\begin{aligned}
g'(x) &= \left(\frac{1}{1+e^{-x}}\right)' = \left((1+e^{-x})^{-1}\right)' = \frac{1}{(1+e^{-x})^2}\cdot e^{-x} \\
&= \frac{1}{(1+e^{-x})}\cdot\frac{e^{-x}}{(1+e^{-x})} \\
&= \frac{1}{(1+e^{-x})}\cdot\frac{1}{(1+e^{x})}
\end{aligned}
$$

Posteriors and the Logistic Function (cont.)

The derivative of the sigmoid function fulfills the nice property:

$$
\begin{aligned}
g'(x) &= \left(\frac{1}{1+e^{-x}}\right)' = \left((1+e^{-x})^{-1}\right)' = \frac{1}{(1+e^{-x})^2} \cdot e^{-x} \\
&= \frac{1}{(1+e^{-x})} \cdot \frac{e^{-x}}{(1+e^{-x})} \\
&= \frac{1}{(1+e^{-x})} \cdot \frac{1}{(1+e^{x})} \\
&= g(x)g(-x)
\end{aligned}
$$

Posteriors and the Logistic Function (cont.)

The derivative of the sigmoid function fulfills the nice property:

$$
\begin{aligned}
g'(x) &= \left(\frac{1}{1+e^{-x}}\right)' = \left((1+e^{-x})^{-1}\right)' = \frac{1}{(1+e^{-x})^2} \cdot e^{-x} \\
&= \frac{1}{(1+e^{-x})} \cdot \frac{e^{-x}}{(1+e^{-x})} \\
&= \frac{1}{(1+e^{-x})} \cdot \frac{1}{(1+e^{x})} \\
&= g(x)g(-x) \\
&= g(x)(1-g(x)) \quad .
\end{aligned}
$$

Posteriors and the Logistic Function (cont.)

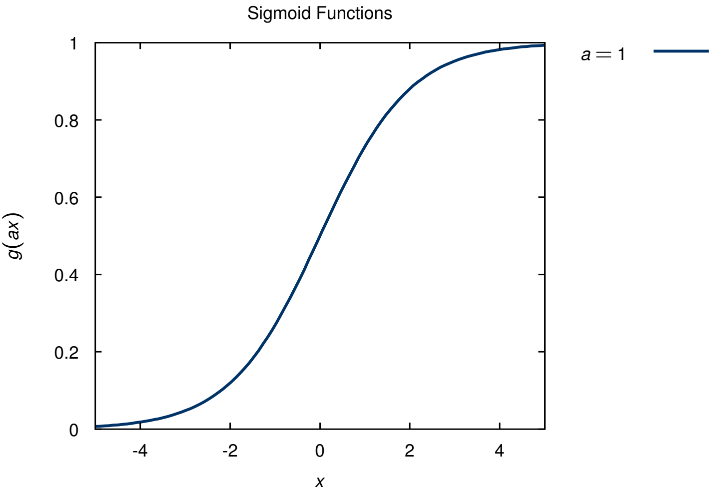


Fig.: Sigmoid function: $g(ax) = 1/(1+e^{-ax})$ for $a = 1, 2, 3, 4$

Posteriors and the Logistic Function (cont.)

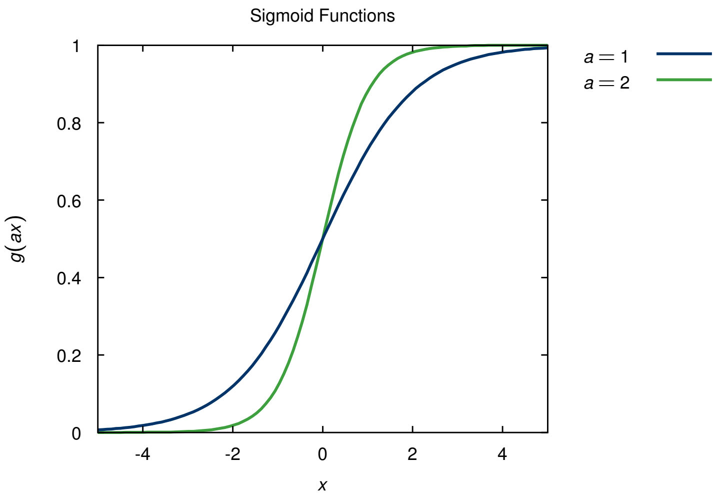


Fig.: Sigmoid function: $g(ax) = 1/(1 + e^{-ax})$ for $a = 1, 2, 3, 4$

Posteriors and the Logistic Function (cont.)

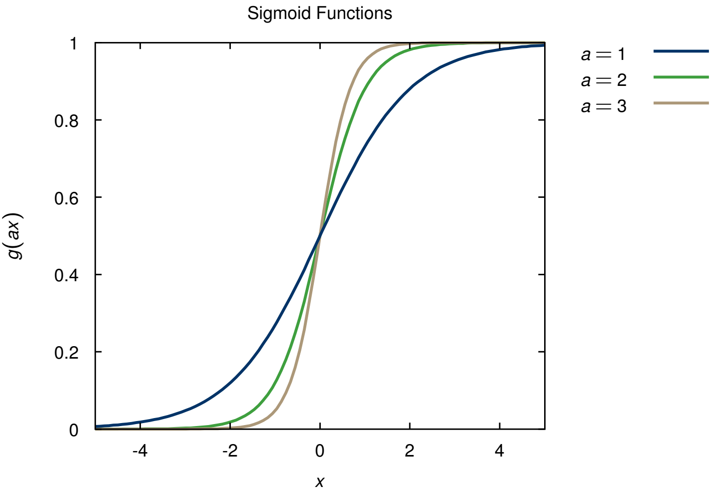


Fig.: Sigmoid function: $g(ax) = 1/(1+e^{-ax})$ for $a = 1, 2, 3, 4$

Posteriors and the Logistic Function (cont.)

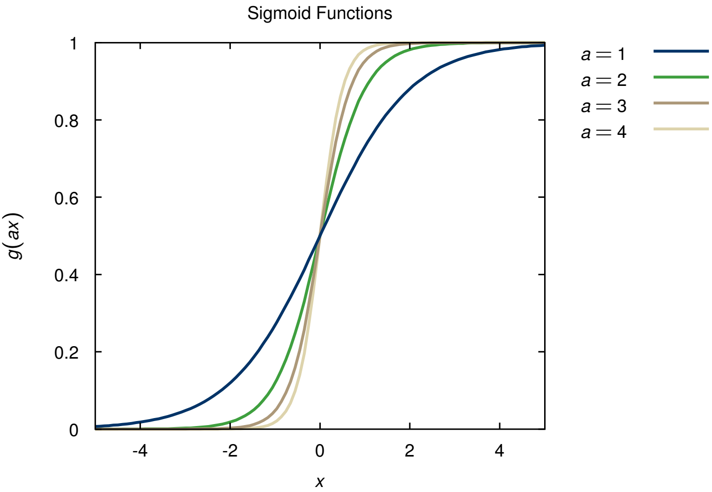


Fig.: Sigmoid function: $g(ax) = 1/(1 + e^{-ax})$ for $a = 1, 2, 3, 4$

Pattern Recognition Lab

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Next Time in

Pattern Recognition

Decision Boundary

The decision boundary $\delta(\boldsymbol{x}) = 0$ (zero level set) in feature space separates the two classes.

Points $\boldsymbol{x}$ on the decision boundary satisfy:

$$p(y = 0|\boldsymbol{x}) \quad = \quad p(y = 1|\boldsymbol{x})$$

and thus

$$\log \frac{p(y = 0|\boldsymbol{x})}{p(y = 1|\boldsymbol{x})} \quad =$$

Decision Boundary

The decision boundary $\delta(\boldsymbol{x}) = 0$ (zero level set) in feature space separates the two classes.

Points $\boldsymbol{x}$ on the decision boundary satisfy:

$$p(y = 0|\boldsymbol{x}) \quad = \quad p(y = 1|\boldsymbol{x})$$

and thus

$$\log \frac{p(y = 0|\boldsymbol{x})}{p(y = 1|\boldsymbol{x})} \quad = \quad \log 1 = 0 \quad .$$

Decision Boundary (cont.)

Lemma

The decision boundary is given by $F(\boldsymbol{x}) = 0$.

Decision Boundary (cont.)

Lemma

The decision boundary is given by $F(\boldsymbol{x}) = 0$.

Proof:

$$\log \frac{p(y=0|\boldsymbol{x})}{p(y=1|\boldsymbol{x})} = F(\boldsymbol{x}) = 0$$

Decision Boundary (cont.)

Lemma

The decision boundary is given by $F(\boldsymbol{x}) = 0$.

Proof:

$$\log \frac{p(y=0|\boldsymbol{x})}{p(y=1|\boldsymbol{x})} = F(\boldsymbol{x}) = 0$$

$$\frac{p(y=0|\boldsymbol{x})}{p(y=1|\boldsymbol{x})} = e^{F(\boldsymbol{x})}$$

Decision Boundary (cont.)

Lemma

The decision boundary is given by $F(\boldsymbol{x}) = 0$.

Proof:

$$\log \frac{p(y=0|\boldsymbol{x})}{p(y=1|\boldsymbol{x})} = F(\boldsymbol{x}) = 0$$

$$\frac{p(y=0|\boldsymbol{x})}{p(y=1|\boldsymbol{x})} = e^{F(\boldsymbol{x})}$$

$$p(y=0|\boldsymbol{x}) = e^{F(\boldsymbol{x})} p(y=1|\boldsymbol{x})$$

Decision Boundary (cont.)

Now we use that the posteriors sum up to one:

$$p(y=0|\boldsymbol{x}) = e^{F(\boldsymbol{x})}\left(1-p(y=0|\boldsymbol{x})\right)$$

Decision Boundary (cont.)

Now we use that the posteriors sum up to one:

$$p(y=0|\boldsymbol{x}) \quad = \quad e^{F(\boldsymbol{x})}\left(1-p(y=0|\boldsymbol{x})\right)$$

$$p(y=0|\boldsymbol{x}) \quad = \quad \frac{e^{F(\boldsymbol{x})}}{1+e^{F(\boldsymbol{x})}}$$

Decision Boundary (cont.)

Now we use that the posteriors sum up to one:

$$
\begin{aligned}
p(y=0|\boldsymbol{x}) &= e^{F(\boldsymbol{x})}\left(1-p(y=0|\boldsymbol{x})\right) \\
p(y=0|\boldsymbol{x}) &= \frac{e^{F(\boldsymbol{x})}}{1+e^{F(\boldsymbol{x})}} \\
p(y=0|\boldsymbol{x}) &= \frac{1}{1+e^{-F(\boldsymbol{x})}}
\end{aligned}
$$

Decision Boundary (cont.)

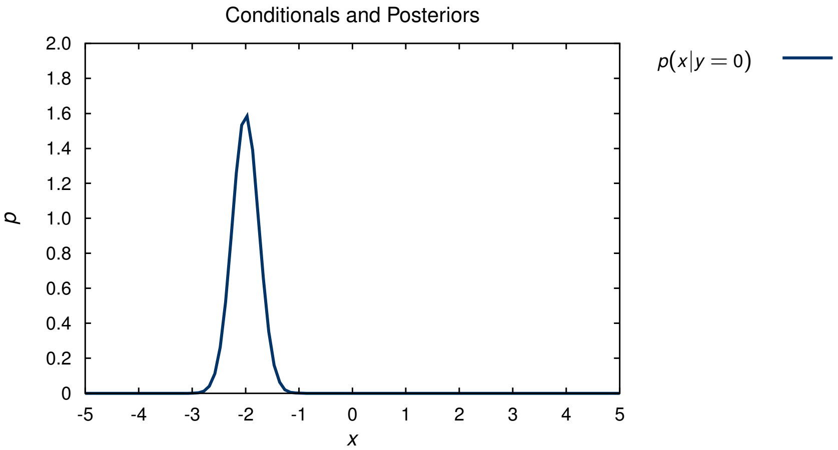


Fig.: Two Gaussians and their posteriors: $\sigma_0 = \sigma_1 = 0.25$, $\mu_0 = -2$, $\mu_1 = 1$

Decision Boundary (cont.)

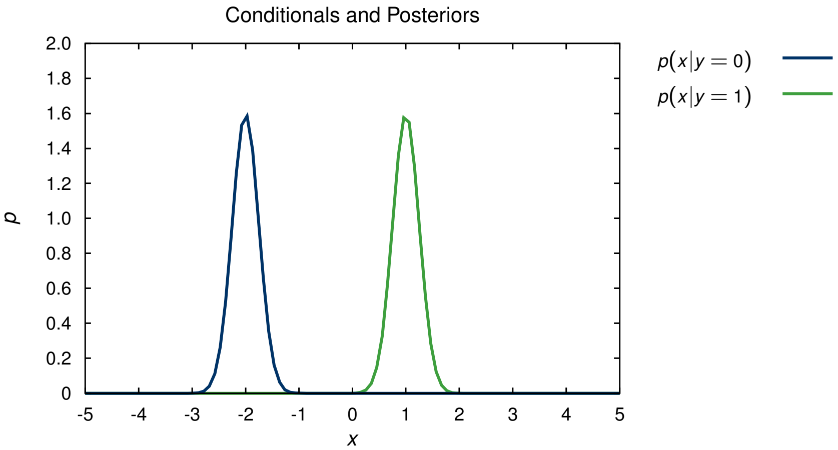


Fig.: Two Gaussians and their posteriors: σ_0=σ_1= 0.25, $\mu_0 = -2$, $\mu_1 = 1$

Decision Boundary (cont.)

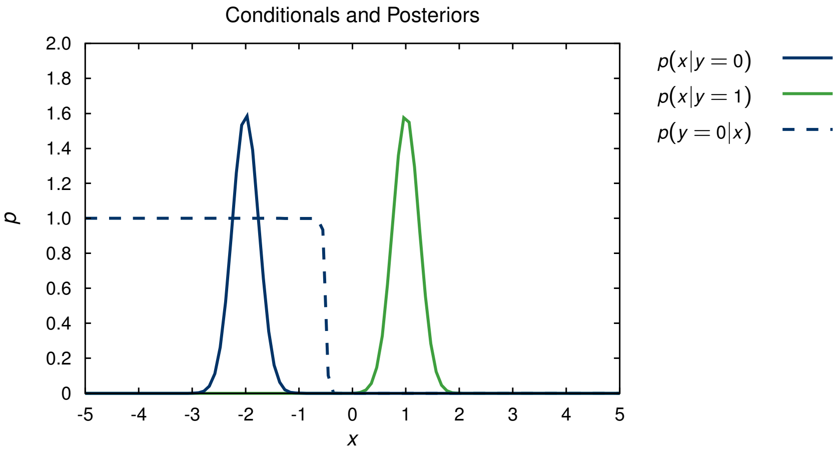


Fig.: Two Gaussians and their posteriors: σ_0=σ_1= 0.25, $\mu_0 = -2$, $\mu_1 = 1$

Decision Boundary (cont.)

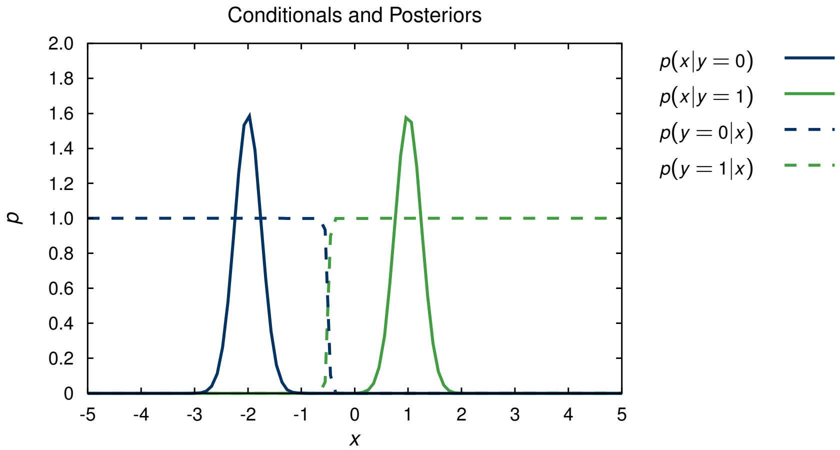


Fig.: Two Gaussians and their posteriors: $\sigma_0 = \sigma_1 = 0.25$, $\mu_0 = -2$, $\mu_1 = 1$

Decision Boundary (cont.)

Example

Let us assume both classes have normally distributed d-dimensional feature vectors:

$$p(\boldsymbol{x}|y) \quad = \quad \frac{1}{\sqrt{\det\left(2\pi\boldsymbol{\Sigma}_y\right)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_y)^T\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_y)}$$

Decision Boundary (cont.)

Example

Let us assume both classes have normally distributed *d*-dimensional feature vectors:

$$p(\boldsymbol{x}|y) = \frac{1}{\sqrt{\det\left(2\pi\boldsymbol{\Sigma}_y\right)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_y)^T\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_y)}$$

Then we can write the posterior of $y = 0$ in terms of a logistic function:

$$p(y = 0|\boldsymbol{x}) = \frac{1}{1 + e^{-F(\boldsymbol{x})}} = \frac{1}{1 + e^{-(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}+\boldsymbol{\alpha}^T\boldsymbol{x}+\alpha_0)}}$$

Decision Boundary (cont.)

Example

Let us assume both classes have normally distributed *d*-dimensional feature vectors:

$$p(\boldsymbol{x}|y) = \frac{1}{\sqrt{\det\left(2\pi\boldsymbol{\Sigma}_y\right)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_y)^T\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_y)}$$

Then we can write the posterior of $y = 0$ in terms of a logistic function:

$$p(y=0|\boldsymbol{x}) = \frac{1}{1+e^{-F(\boldsymbol{x})}} = \frac{1}{1+e^{-(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}+\boldsymbol{\alpha}^T\boldsymbol{x}+\alpha_0)}}$$

$$F(\boldsymbol{x}) = \log\frac{p(y=0|\boldsymbol{x})}{p(y=1|\boldsymbol{x})} = \log\frac{p(y=0)p(\boldsymbol{x}|y=0)}{p(y=1)p(\boldsymbol{x}|y=1)}$$

Decision Boundary (cont.)

Example cont.

$$F(\boldsymbol{x}) \quad = \quad \log \frac{p(y=0)}{p(y=1)} + \log \frac{\frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_0)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)}}{\frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_1)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}}$$

Decision Boundary (cont.)

Example cont.

$$F(\boldsymbol{x}) \quad = \quad \log \frac{p(y=0)}{p(y=1)} + \log \frac{\frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_0)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)}}{\frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_1)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}}$$

This function has the constant component:

$$c \quad = \quad \log \frac{p(y=0)}{p(y=1)} + \frac{1}{2} \log \frac{\det(2\pi\boldsymbol{\Sigma}_1)}{\det(2\pi\boldsymbol{\Sigma}_0)}$$

Decision Boundary (cont.)

Example cont.

$$F(\boldsymbol{x}) \quad = \quad \log \frac{p(y=0)}{p(y=1)} + \log \frac{\frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_0)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)}}{\frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_1)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}}$$

This function has the constant component:

$$c \quad = \quad \log \frac{p(y=0)}{p(y=1)} + \frac{1}{2} \log \frac{\det(2\pi\boldsymbol{\Sigma}_1)}{\det(2\pi\boldsymbol{\Sigma}_0)}$$

We observe:

- Priors imply a constant offset of the decision boundary.
- If priors and covariance matrices of both classes are identical, this offset is $c = 0$.

Decision Boundary (cont.)

Example cont.

Furthermore we have:

$$\log \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)}}{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}} \quad =$$

Decision Boundary (cont.)

Example cont.

Furthermore we have:

$$\log \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)}}{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}} =$$

$$= \frac{1}{2}\left((\boldsymbol{x}-\boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) - (\boldsymbol{x}-\boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\right)$$

Decision Boundary (cont.)

Example cont.

Furthermore we have:

$$\log \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)}}{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}} =$$

$$= \frac{1}{2}\left((\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) - (\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\right)$$

$$= \frac{1}{2}\left(\boldsymbol{x}^T(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1})\boldsymbol{x} - 2(\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1})\boldsymbol{x} + \right.$$
$$\left. + \boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right)$$

Decision Boundary (cont.)

Example cont.

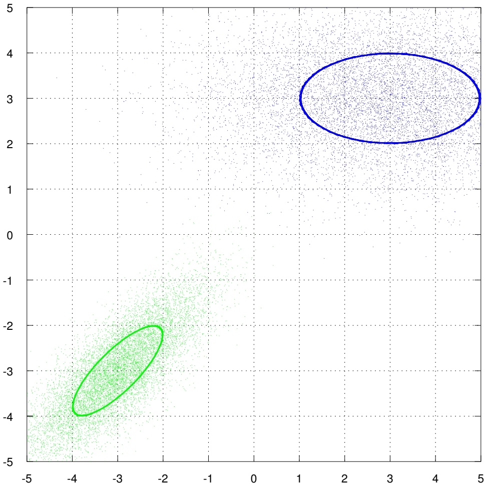
Now we have:

$$\boldsymbol{A} = \frac{1}{2}(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1})$$

$$\boldsymbol{\alpha}^T = \boldsymbol{\mu}_0^T \boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}_1^{-1}$$

$$\alpha_0 = \log \frac{p(y=0)}{p(y=1)} + \frac{1}{2}\left(\log \frac{\det(2\pi\boldsymbol{\Sigma}_1)}{\det(2\pi\boldsymbol{\Sigma}_0)} + \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0\right)$$

Decision Boundary (cont.)

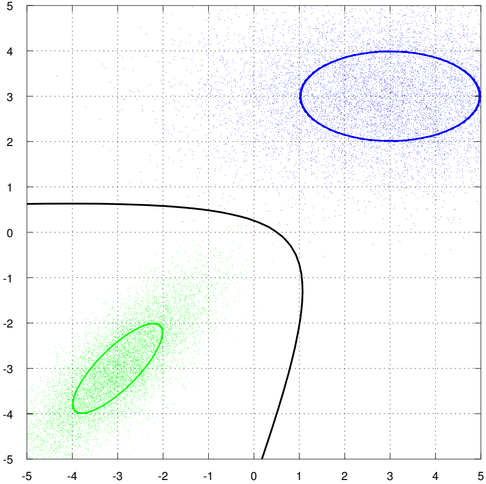


$p(y = 0) = 0.5$

$p(y = 1) = 0.5$

Fig.: Two Gaussian sample sets and the decision boundary

Decision Boundary (cont.)

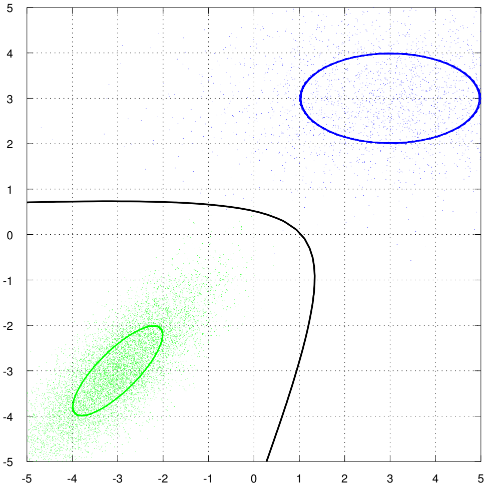


$p(y = 0) = 0.5$

$p(y = 1) = 0.5$

Fig.: Two Gaussian sample sets and the decision boundary

Decision Boundary (cont.)

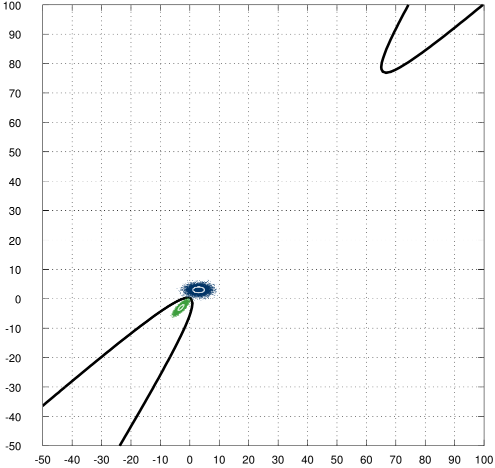


$p(y = 0) = 0.8$

$p(y = 1) = 0.2$

Fig.: Two Gaussian sample sets and the decision boundary

Decision Boundary (cont.)



$p(y = 0) = 0.5$

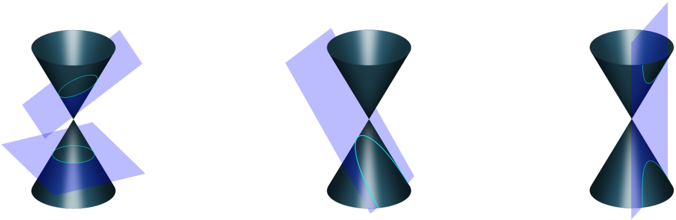
$p(y = 1) = 0.5$

Fig.: Two Gaussian sample sets and the decision boundary

Decision Boundary (cont.)

Quadratic polynomials in the 2 variables x_1 and x_2

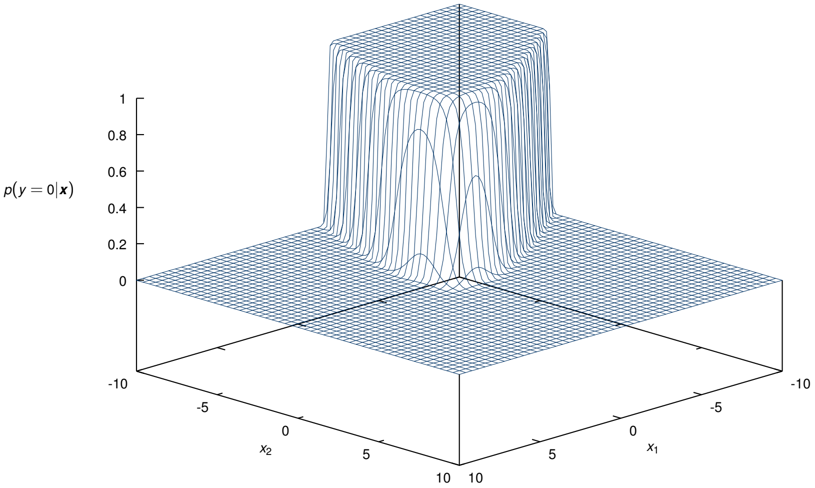
$$\begin{aligned} F(\boldsymbol{x}) &= \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\alpha}^T \boldsymbol{x} + \alpha_0 \\ &= a x_1^2 + b x_1 x_2 + c x_2^2 + d x_1 + e x_2 + f \stackrel{!}{=} 0 \end{aligned}$$



(a) circles and ellipses (b) parabolas (c) hyperbolas

Decision Boundary (cont.)

Posterior probability



Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Next Time in

Pattern Recognition

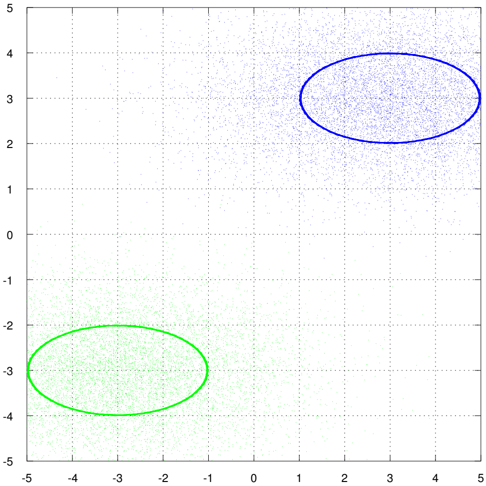
Decision Boundary in Distributions with Equal Dispersion

Example cont.

If both classes share the same covariances i. e. $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1$, then the argument of the sigmoid function is linear in the components of $\boldsymbol{x}$.

$$\begin{aligned}
\boldsymbol{A} &= 0 \\
\boldsymbol{\alpha}^T &= (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} \\
\alpha_0 &= \log \frac{p(y=0)}{p(y=1)} + \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)
\end{aligned}$$

Decision Boundary in Distributions with Equal Dispersion (cont.)

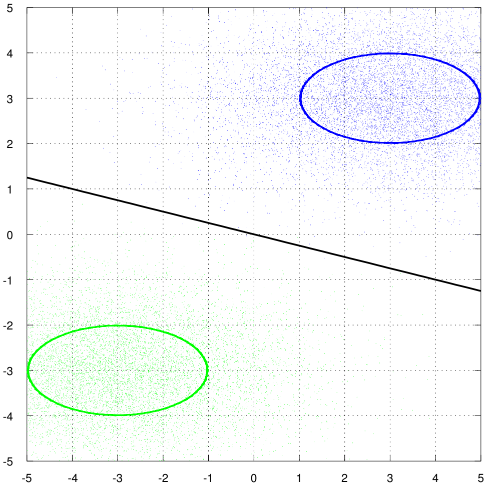


$$p(y=0) = 0.5$$

$$p(y=1) = 0.5$$

Fig.: Identical covariances lead to linear decision boundary

Decision Boundary in Distributions with Equal Dispersion (cont.)



$$p(y = 0) = 0.5$$

$$p(y = 1) = 0.5$$

Fig.: Identical covariances lead to linear decision boundary

Decision Boundary in Distributions with Equal Dispersion (cont.)

Note:

- If the class conditionals are Gaussians and share the same covariance, the argument of the exponential function is affine in $\boldsymbol{x}$.
- This result is even true for a more general family of pdfs and not limited to Gaussians.

Decision Boundary in Distributions with Equal Dispersion (cont.)

Definition

The *exponential family* is a class of pdf's that can be written in the following canonical form

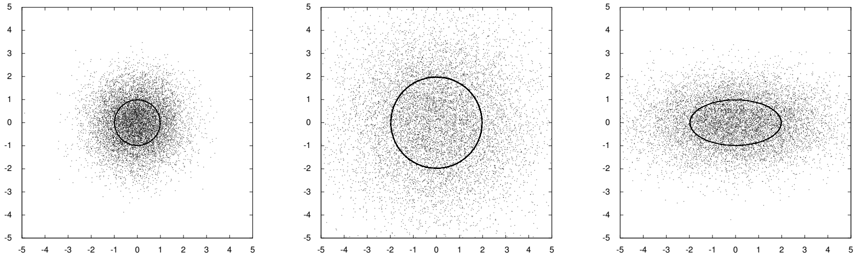
$$p(\boldsymbol{x}; \boldsymbol{\theta}, \phi) \quad = \quad e^{\frac{\boldsymbol{\theta}^T \cdot \boldsymbol{x} - b(\boldsymbol{\theta})}{a(\phi)} + c(\boldsymbol{x}, \phi)}$$

where $\boldsymbol{\theta} \in \mathbb{R}^d$ is the *location parameter vector*, ϕ the *dispersion parameter*.

Exponential Family

Gaussian Probability Density Function

$$\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_y)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$



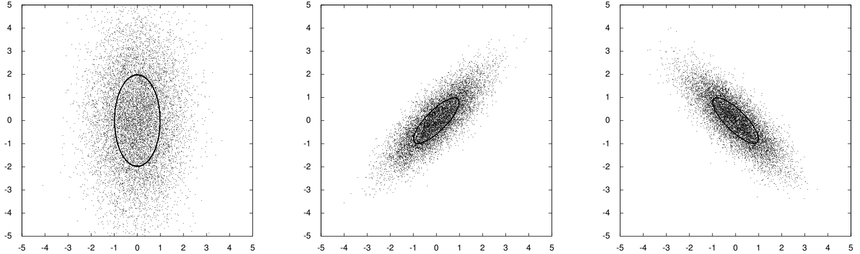
$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$ $\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$

Fig.: Gaussian probability density functions with $\boldsymbol{\mu} = (0,0)^T$

Exponential Family

Gaussian Probability Density Function (cont.)

$$\mathscr{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{\det\left(2\pi\boldsymbol{\Sigma}_y\right)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$



$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$ $\quad$ $\boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{pmatrix}$ $\quad$ $\boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & -0.8 \\ -0.8 & 1.0 \end{pmatrix}$

Fig.: Gaussian probability density functions with $\boldsymbol{\mu} = (0,0)^T$

Exponential Family (cont.)

Exponential Probability Density Function

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

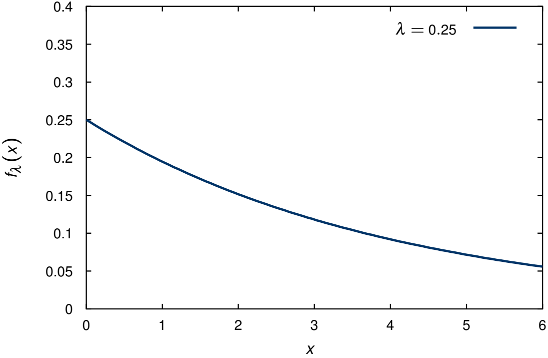


Fig.: Exponential probability density functions

Exponential Family (cont.)

Exponential Probability Density Function

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

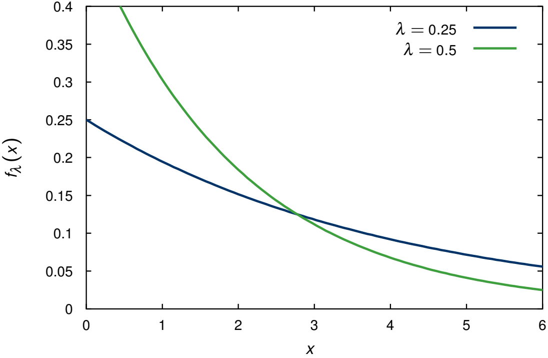


Fig.: Exponential probability density functions

Exponential Family (cont.)

Exponential Probability Density Function

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

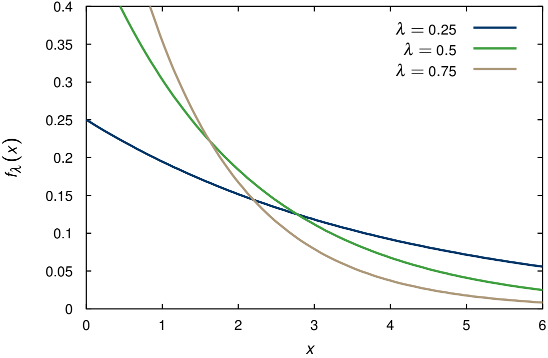


Fig.: Exponential probability density functions

Exponential Family (cont.)

Exponential Probability Density Function

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

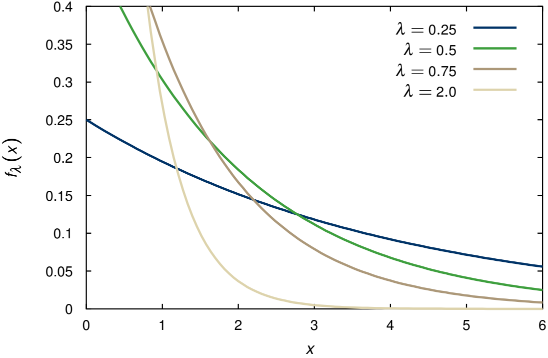


Fig.: Exponential probability density functions

Exponential Family (cont.)

Binomial Probability Mass Function

$$B(k;p,n) = \binom{n}{k} p^k (1-p)^{n-k}$$

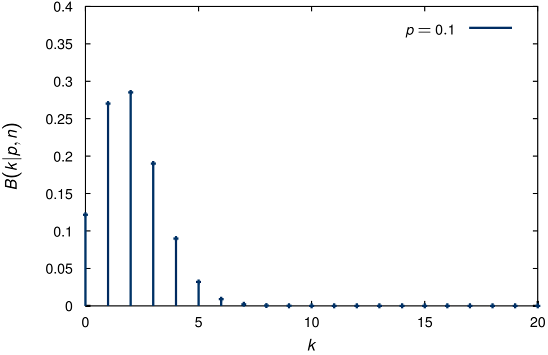


Fig.: Binomial probability mass functions for $n = 20$

Exponential Family (cont.)

Binomial Probability Mass Function

$$B(k;p,n) = \binom{n}{k} p^k (1-p)^{n-k}$$

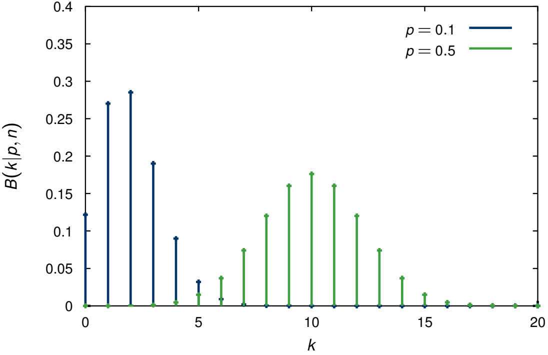


Fig.: Binomial probability mass functions for $n = 20$

Exponential Family (cont.)

Binomial Probability Mass Function

$$B(k; p, n) = \binom{n}{k} p^k (1-p)^{n-k}$$

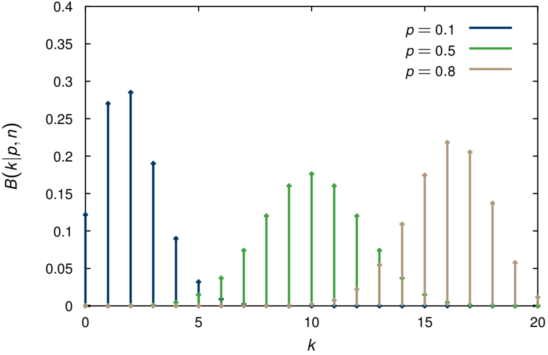


Fig.: Binomial probability mass functions for $n = 20$

Exponential Family (cont.)

Poisson Probability Mass Function

$$P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

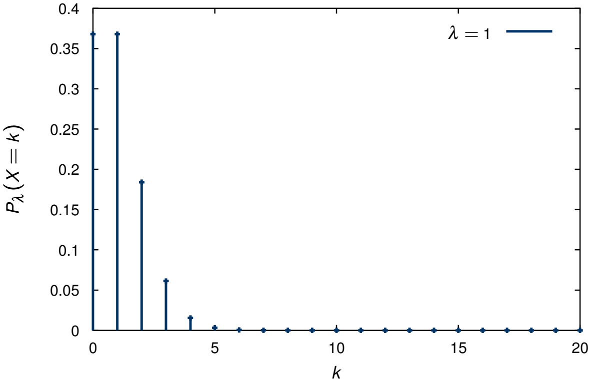


Fig.: Poisson probability mass functions

Exponential Family (cont.)

Poisson Probability Mass Function

$$P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

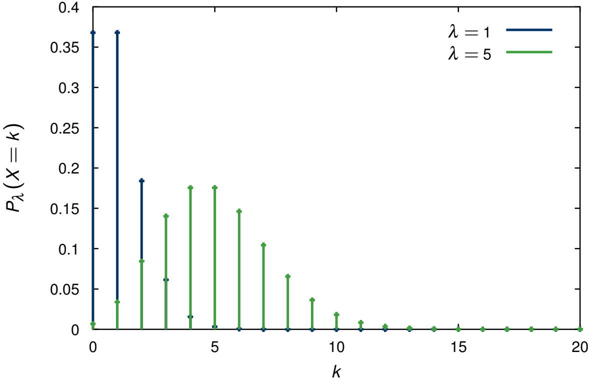


Fig.: Poisson probability mass functions

Exponential Family (cont.)

Poisson Probability Mass Function

$$P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

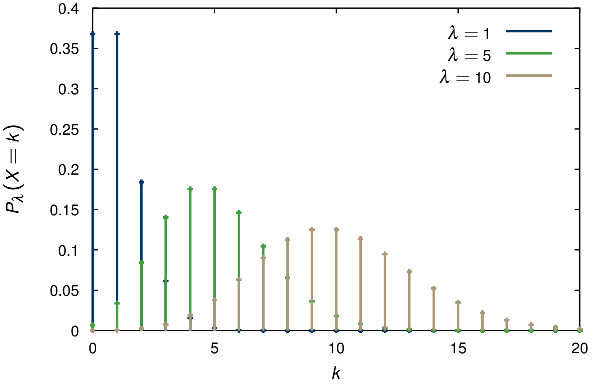


Fig.: Poisson probability mass functions

Exponential Family (cont.)

Hypergeometric Probability Mass Function

$$h(k; N, M, n) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

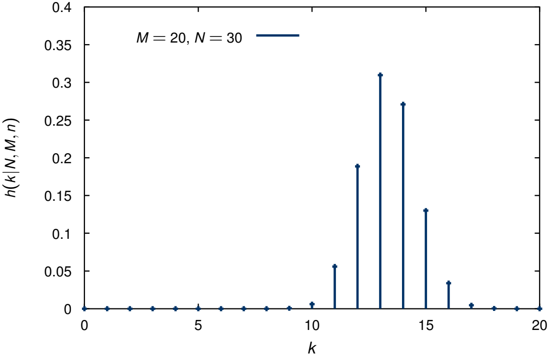


Fig.: Hypergeometric probability mass functions

Exponential Family (cont.)

Hypergeometric Probability Mass Function

$$h(k; N, M, n) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

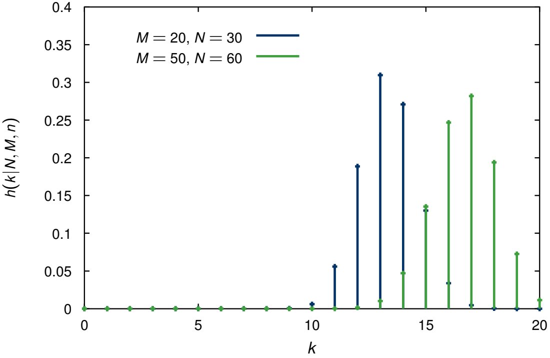


Fig.: Hypergeometric probability mass functions

Exponential Family (cont.)

Hypergeometric Probability Mass Function

$$h(k;N,M,n) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

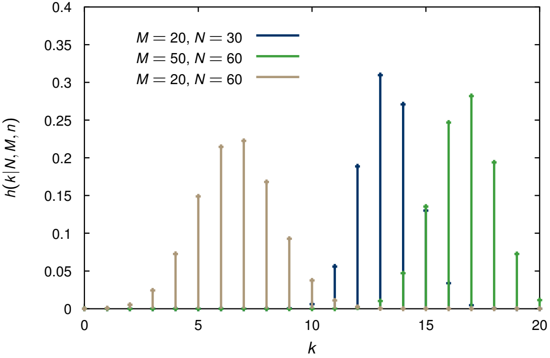


Fig.: Hypergeometric probability mass functions

Decision Boundary (cont.)

Lemma

If all class-conditional densities are members of the same exponential family of probability density functions with equal dispersion ϕ, the decision boundary $F(\boldsymbol{x}) = 0$ is linear in the components of $\boldsymbol{x}$.

Lessons Learned

- Posteriors can be rewritten in terms of a logistic function.

Lessons Learned

- Posteriors can be rewritten in terms of a logistic function.
- Given the decision boundary $F(\boldsymbol{x}) = 0$, we can write down the posterior $p(y|\boldsymbol{x})$ right away.

Lessons Learned

- Posteriors can be rewritten in terms of a logistic function.
- Given the decision boundary $F(\boldsymbol{x}) = 0$, we can write down the posterior $p(y|\boldsymbol{x})$ right away.
- Decision boundary for normally distributed feature vectors for each class is a quadratic function.

Lessons Learned

- Posteriors can be rewritten in terms of a logistic function.
- Given the decision boundary $F(\boldsymbol{x}) = 0$, we can write down the posterior $p(y|\boldsymbol{x})$ right away.
- Decision boundary for normally distributed feature vectors for each class is a quadratic function.
- If Gaussians share the same covariances, the decision boundary is a linear function.

Pattern Recognition Lab

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Next Time in

Pattern Recognition

Further Readings

- T. Hastie, R. Tibshirani, and J. Friedman: The Elements of Statistical Learning – Data Mining, Inference, and Prediction, 2nd edition, Springer, New York, 2009.
- David W. Hosmer, Stanley Lemeshow: Applied Logistic Regression, 2nd Edition, John Wiley & Sons, Hoboken, 2000.

Comprehensive Questions

- How can we model the posterior probabilities?

Comprehensive Questions

- How can we model the posterior probabilities?

- Formulate the criterion for the decision boundary!

Comprehensive Questions

- How can we model the posterior probabilities?
- Formulate the criterion for the decision boundary!
- Describe the shape of the decision boundary for a Gaussian with different and same class covariances!

Comprehensive Questions

- How can we model the posterior probabilities?
- Formulate the criterion for the decision boundary!
- Describe the shape of the decision boundary for a Gaussian with different and same class covariances!
- What effect does a change of the priors have on the decision boundary?