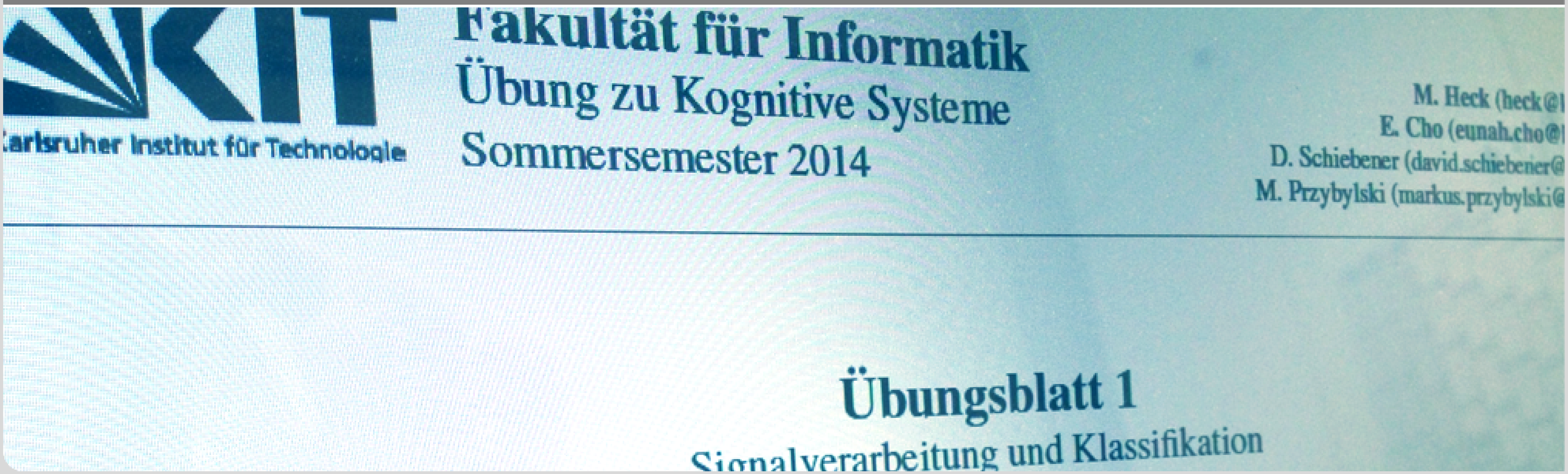
Kognitive Systeme – Übung 2

12.06.2017 – Klassifikation und Spracherkennung

Matthias Sperber, Thai Son Nguyen

KIT, Institute for Anthropomatics and Robotics, Department of Informatics, Interactive Systems Laboratories



- Exercise 1: Naive Bayes Classifier
- Exercise 2: Probability of Errors
- Exercise 3: K Nearest Neighbours
- Exercise 4: Perceptron
- Exercise 5: Perceptron Learning

Exercise 1a

- Ask you to produce N random numbers

- When n = 10, we have for examples:
  - 5.951; 6.196; 6.409; 4.908; 4.876; 5.590; 6.102; 4.915; 4.818; 6.136;

Exercise 1b

- Gaussian (or Normal) distribution:
  - Has the probability density as:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where:

- μ is mean or expectation of the distribution (and also its median and mode).
- σ is standard deviation
- σ^2 is variance

  - The maximum likelihood estimators of the mean and the variance:

$$\hat{\mu}_n = \frac{1}{n}\sum_{j=1}^{n} x_j$$

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{j=1}^{n} (x_j - \hat{\mu})^2$$

Exercise 1b

- Expected **mean** and **variance** for the distribution from the 10 samples:

$$
\begin{aligned}
\mu &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \frac{1}{10}\cdot(5.951+6.196+6.409+4.908+4.876+5.590+6.102 \\
&\quad +4.915+4.818+6.136) \\
&\approx 5.590
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 &= \sum_{i=1}^{n}(x_i-\mu)^2\cdot\frac{1}{n} \\
&= (5.951-5.590)^2\cdot\frac{1}{10}+(6.196-5.590)^2\cdot\frac{1}{10}+(6.409-5.590)^2\cdot\frac{1}{10}+ \\
&\quad (4.908-5.590)^2\cdot\frac{1}{10}+(4.876-5.590)^2\cdot\frac{1}{10}+(5.590-5.590)^2\cdot\frac{1}{10}+ \\
&\quad (6.102-5.590)^2\cdot\frac{1}{10}+(4.915-5.590)^2\cdot\frac{1}{10}+(4.818-5.590)^2\cdot\frac{1}{10}+ \\
&\quad (6.136-5.590)^2\cdot\frac{1}{10} \\
&\approx 0.645 \\
\sigma &\approx 0.803
\end{aligned}
$$

Exercise 1c

- We have a Gaussian distribution from 1b)
  - p(x|ω1) ~ N (5.590; 0.645)
- Assume we have another distribution:
  - p(x|ω2) ~ N (5.5; 1) ~ μ = 5.5 and σ^2 = 1

- Given a new random number **x**, we want to estimate which distribution (ω1 or ω2) the number likely belongs to
  - Bayes Classifier: assign x to the distribution (or class) which has the highest **posterior probability** $p(\omega_c|x)$ (Maximum a posteriori - MAP)
  - **Posterior probability** in Bayes theorem is defined as:

$$P(\omega_c|x) = \frac{p(x|\omega_c)P(\omega_c)}{p(x)}$$

Exercise 1c

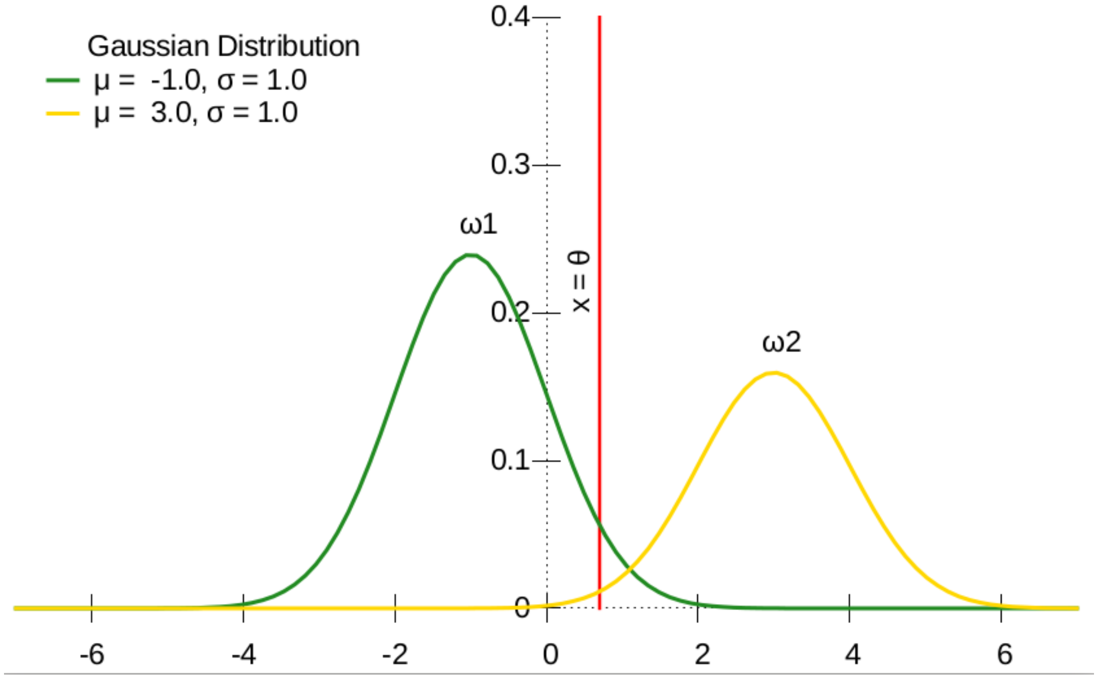
- x = 5.805
- **p(ω1|x) <> p(ω2|x)?**
- p(ω1|x).p(ω1) <> p(ω2|x).p(ω2)
  - p(x) is skipped
  - p(x|ω1),p(x|ω2) are known
  - p(ω1),p(ω2) are also known

$$
\begin{aligned}
P(\omega_1|x) &= \frac{p(x|\omega_1)P(\omega_1)}{p(x)} \\
p(x|\omega_1)P(\omega_1) &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{(-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2)} P(\omega_1) \\
&= \frac{1}{\sqrt{2\pi} \cdot 0.803} e^{(-\frac{1}{2}(\frac{5.805-5.590}{0.803})^2)} \cdot 0.3 \\
&\approx 0.497 \cdot e^{-0.036} \cdot 0.3 \approx 0.144
\end{aligned}
$$

$$
\begin{aligned}
P(\omega_2|x) &= \frac{p(x|\omega_2)P(\omega_2)}{p(x)} \\
p(x|\omega_2)P(\omega_2) &= \frac{1}{\sqrt{2\pi}\sigma_2} e^{(-\frac{1}{2}(\frac{x-\mu_1}{\sigma_2})^2)} P(\omega_2) \\
&= \frac{1}{\sqrt{2\pi} \cdot 1} e^{(-\frac{1}{2}(\frac{5.805-5.5}{1})^2)} \cdot 0.7 \\
&\approx 0.399 \cdot e^{-0.047} \cdot 0.7 \approx 0.293
\end{aligned}
$$

Exercise 2

- Given two Gaussian distributions ω1 and ω2 of univariate x
- Can we decide which class given x and a threshold θ?
  - e.g. $x < \theta \Rightarrow \omega_1$ and $x > \theta \Rightarrow \omega_2$
  - How to estimate the classification errors for different values of θ?



Exercise 2a

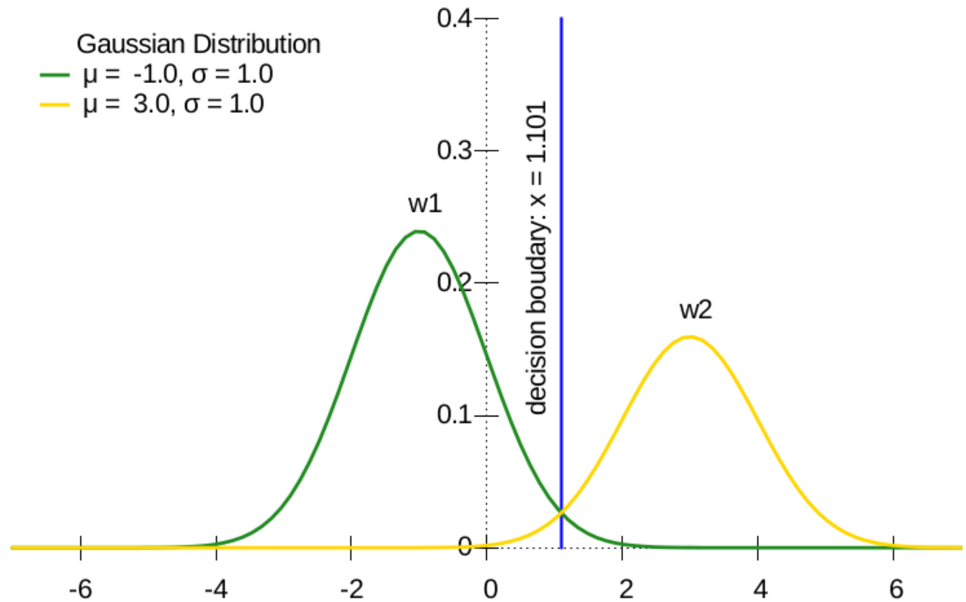
- We estimate the misclassification using **probability of errors** (*the probability of making a wrong decision*):
  - (x ∈ ω1 and x > θ) + (x ∈ ω2 and x < θ)
  - Use probability density function (PDF) but need to scale with class prior

$$P_{Fehler}(\theta) = \int_{\theta}^{\infty} p(x|\omega_1)P(\omega_1)dx + \int_{-\infty}^{\theta} p(x|\omega_2)P(\omega_2)dx$$

Exercise 2b

- Find the θ_{opt} which minimizes $P_{error}(\theta)$
- $P_{error}(\theta)$ is minimized when **p(ω1|θ).p(ω1) = p(ω2|θ).p(ω2)** (*You find the proof yourself!*)

$$
\begin{aligned}
P(\omega_1)p(x|\omega_1) &\overset{!}{=} P(\omega_2)p(x|\omega_2) \\
\frac{3}{5}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\theta_{opt}+1)^2}{2}} &= \frac{2}{5}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\theta_{opt}-3)^2}{2}} \\
3e^{-\frac{(\theta_{opt}+1)^2}{2}} &= 2e^{-\frac{(\theta_{opt}-3)^2}{2}} \\
\frac{3}{2} &= e^{\frac{(\theta_{opt}+1)^2-(\theta_{opt}-3)^2}{2}} \\
1.5 &= e^{4\theta_{opt}-4} \\
ln1.5 &= 4\theta_{opt}-4 \\
\theta_{opt} &= 1+\frac{ln1.5}{4} \approx 1.101
\end{aligned}
$$



Exercise 2c

- The probability of error when $\theta = \theta_{opt}$:

$$
\begin{aligned}
P_{error}(\theta) &= \int_{\theta}^{\infty} p(x|\omega_1)P(\omega_1)dx + \int_{-\infty}^{\theta} p(x|\omega_2)P(\omega_2)dx \\
&= \frac{3}{5}\int_{\theta}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}}dx + \frac{2}{5}\int_{-\infty}^{\theta} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}}dx \\
&= \frac{3}{5}\int_{1.1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}}dx + \frac{2}{5}\int_{-\infty}^{1.1} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}}dx \\
&= \frac{3}{5}\cdot\frac{1}{\sqrt{2\pi}}\int_{1.1+1}^{\infty} e^{-\frac{x^2}{2}}dx + \frac{2}{5}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{1.1-3} e^{-\frac{x^2}{2}}dx \\
&= \frac{3}{5}\cdot(1-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{2.1} e^{-\frac{x^2}{2}}dx) + \frac{2}{5}\cdot(1-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{1.9} e^{-\frac{x^2}{2}}dx) \\
&= \frac{3}{5}\cdot(1-\Phi(2.1)) + \frac{2}{5}\cdot(1-\Phi(1.9)) \\
&\approx \frac{1}{5}\cdot(1-0.98) + \frac{4}{5}\cdot(1-0.97) \\
&= \frac{3}{5}\cdot 0.02 + \frac{2}{5}\cdot 0.03 \\
&\approx 0.022
\end{aligned}
$$

Exercise 2d

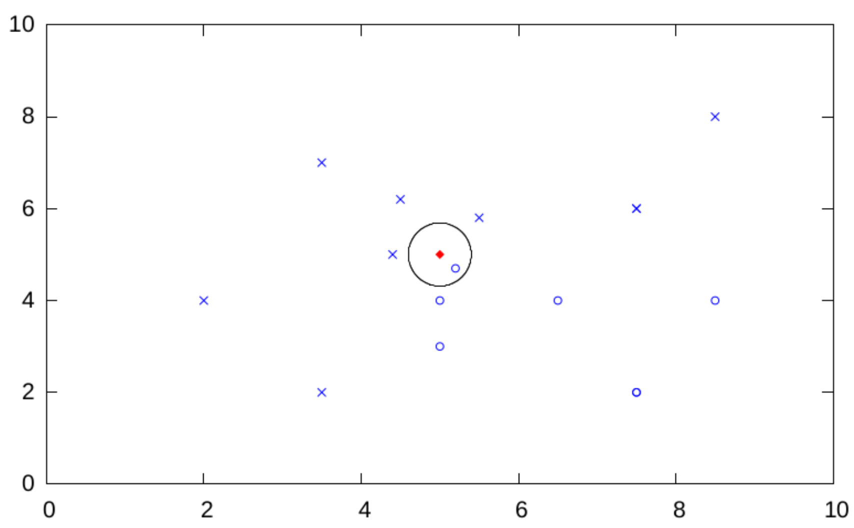
- If the actual error rate is higher than calculated in part 2c:
  - The probability density function is only an estimate that corresponds to the real distribution (as close as possible, but does not perfectly represent them)
  - Too little or distorted measurement data.

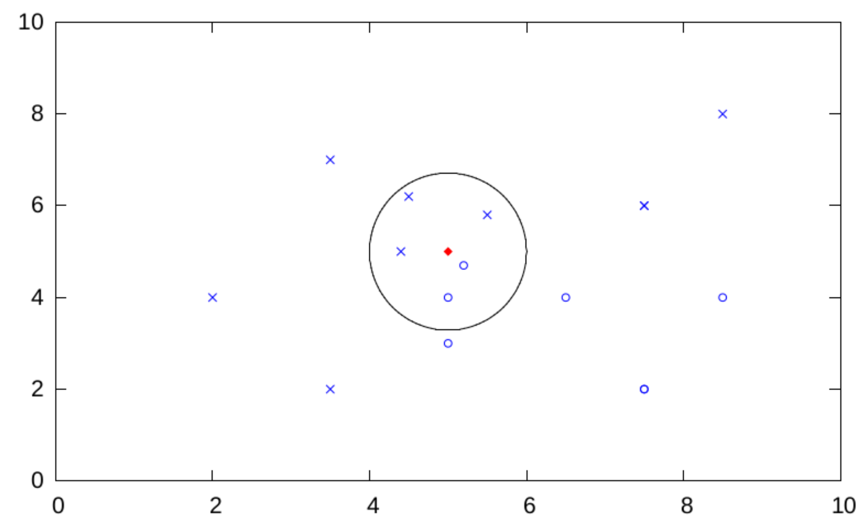
Exercise 3

- K-Nearest-Neighbors classification
  - Nonparametric which is different from Gaussians
  - Simple, almost no learning step (except the selection of k)

Exercise 3

- K-Nearest-Neighbors algorithm: *Find k closest samples (using some distance measure) and decide the class with the largest samples*
  - Using **Euclidean distance**
  - K = 1 => **Circle**, K = 5 => **Cross**





Exercise 4

- Recap:
  - Single perception and linear separability
  - A **single-layer perceptron** is a function of n-dimentional input $(x_1, x_2.., x_n)$

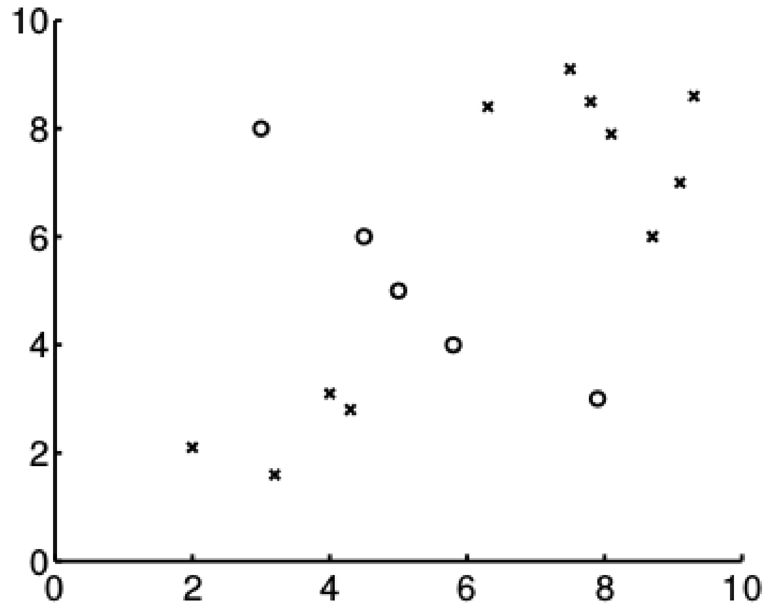
$$g(x) = w_0 + \sum_{i=1}^{n} w_i x_i$$

    - Used to classify patterns said to be linearly separable
  - For 2-dimentional input, the perceptron can be drawn by a line in Cartesian coordinate system

$$g(x) = w_1 x_1 + w_2 x_2 + b.$$

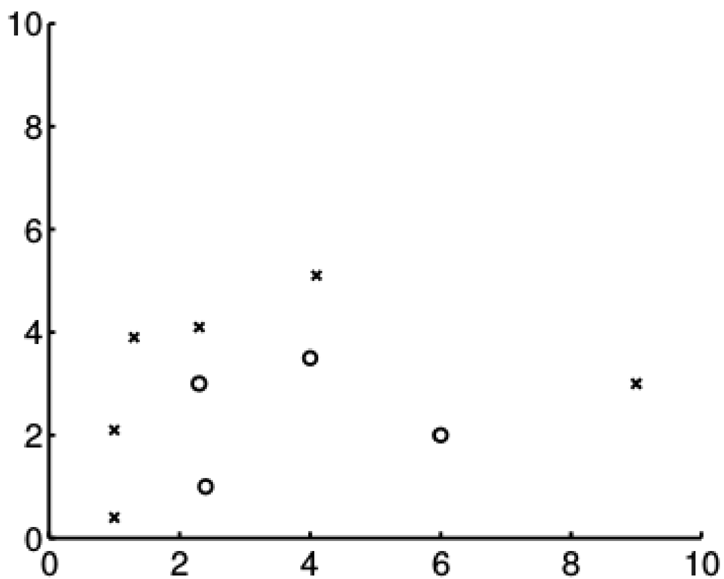
Exercise 4-i

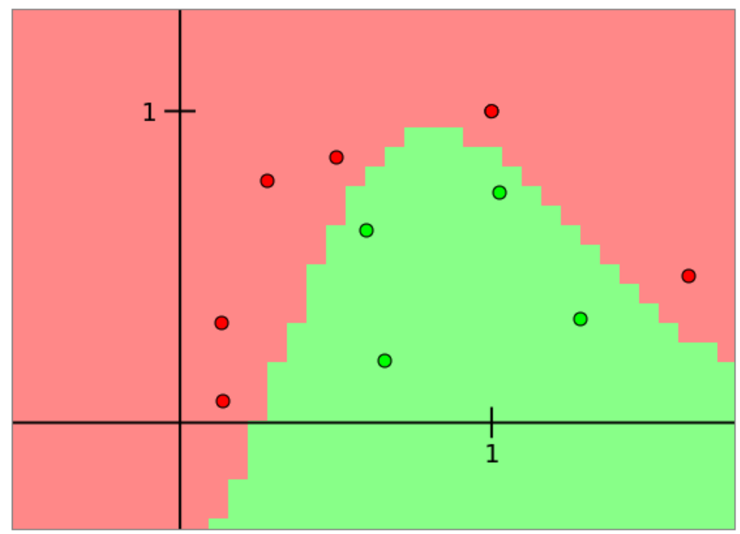
- Two classes are not linearly separable



Exercise 4-ii+iii

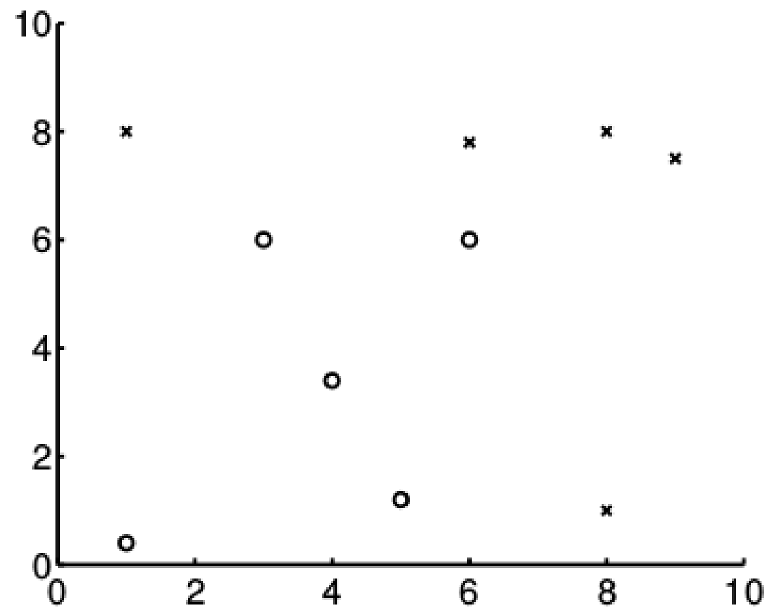
- Two classes are not separated by a single perceptron
- But can be separated by multi-layer perceptron due to **non-linear activiation** functions
  - See a MLP in https://lecture-demo.ira.uka.de/neural-network-demo





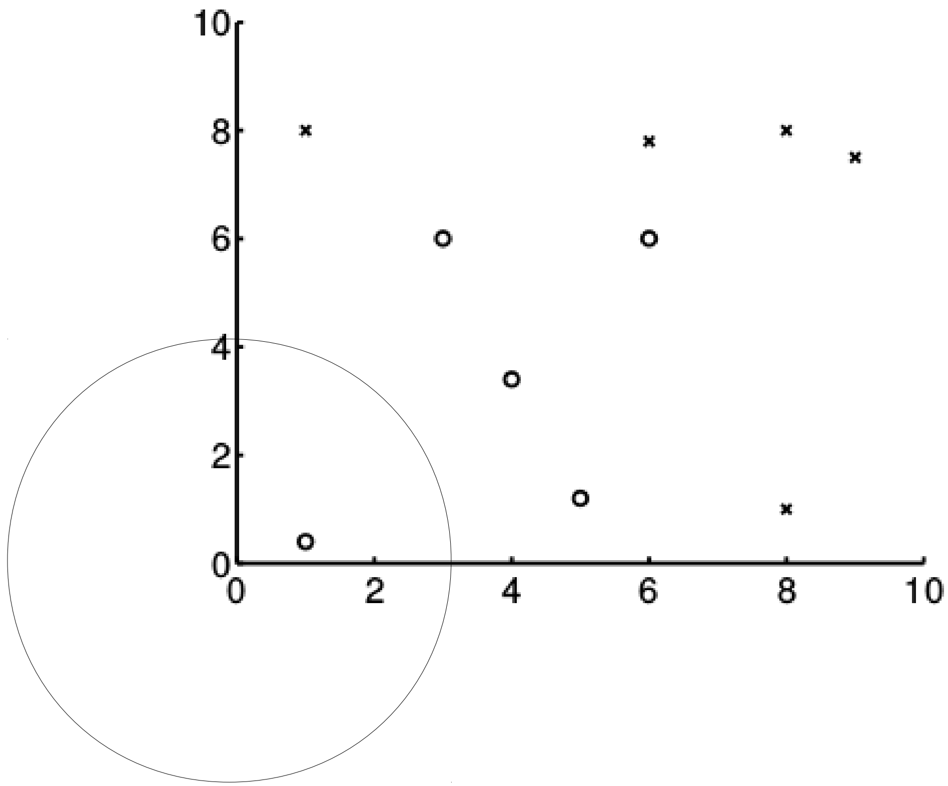
Exercise 4-iv

- Two classes are not linearly separable



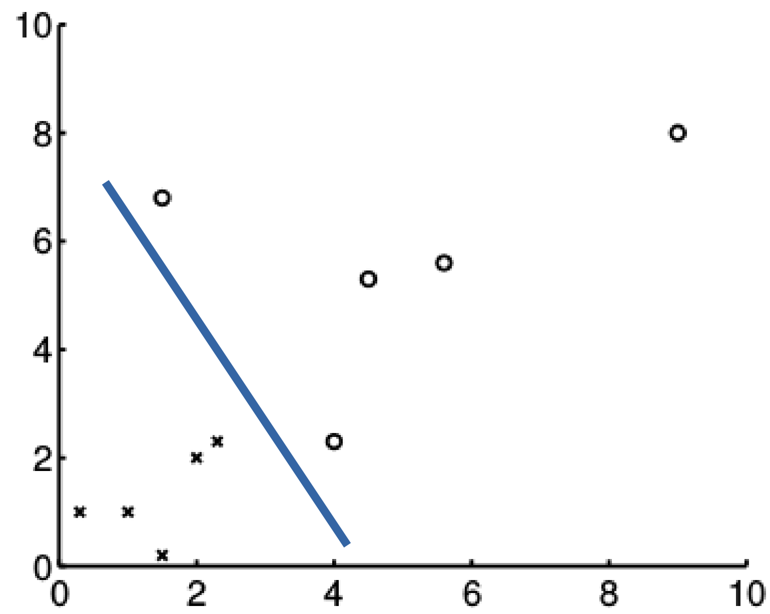
Exercise 4-v

- In **Polar** coordinate system, **a line** includes points having the angle but different distance => will be **a circle** with the center (0, 0) in **Cartesian** coordinate system
  - Not separable



Exercise 4-vi

- Two classes are separable



Exercise 5

- We want to train a perceptron to classify inputs correctly
- Use **Perceptron Learning Algorithm** to find the desired parametters

Require: positive training patterns $\mathcal{P}$ and a negative training examples $\mathcal{N}$
Ensure: if exists, a perceptron is learned that classifies all patterns accurately

initialize weight vector $\vec{w}$ and bias weight w_0 arbitrarily
while exist misclassified pattern $\vec{x} \in \mathcal{P} \cup \mathcal{N}$ **do**
  if $\vec{x} \in \mathcal{P}$ **then**
    $\vec{w} \leftarrow \vec{w} + \vec{x}$
    $w_0 \leftarrow w_0 + 1$
  else
    $\vec{w} \leftarrow \vec{w} - \vec{x}$
    $w_0 \leftarrow w_0 - 1$
  end if
end while
return $\vec{w}$ and w_0

The algorithm is taken from
http://ml.informatik.uni-freiburg.de/_media/documents/teaching/ss09/ml/perceptrons.pdf

Exercise 5

- **Perceptron Convergence Theorem**
  - For any data set which is linearly separable, the perceptron learning rule is guaranteed to find a solution in a finite number of iterations
  - *Find the proof yourself!*

Exercise 5

- The exercise acutally asks you to perform **perceptron learning**
- **Iteration 1:**

1. $\vec{w}_0 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}; b_0 = 0 \implies g_0(x) = x_1 - x_2$

$$g_0(\vec{a}_1) = -1.2 \overset{!}{\implies} B \text{ (Fehlklassifikation)}$$
$$g_0(\vec{a}_2) = -0.6 \overset{!}{\implies} B \text{ (Fehlklassifikation)}$$
$$g_0(\vec{b}_1) = -1.3 \implies B$$
$$g_0(\vec{b}_2) = -1.9 \implies B$$

Damit sind zwei Punkte aus Klasse A falsch klassifiziert, $E_0 = \{\vec{a}_1, \vec{a}_2\}$.

Daraus ergibt sich nach dem ersten Iterationsschritt:

$$\begin{pmatrix} \vec{w}_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - (-1)\vec{a}_1 - (-1)\vec{a}_2$$
$$= \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - (-1)\begin{pmatrix} -0.4 \\ 0.8 \\ 1 \end{pmatrix} - (-1)\begin{pmatrix} -0.8 \\ -0.2 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.2 \\ -0.4 \\ 2 \end{pmatrix}$$
$$\implies g_1(x) = -0.2x_1 - 0.4x_2 + 2$$

Geradengleichung für die Zeichnung: $x_2 = -0.5x_1 + 5$

Exercise 5

- **Iteration 2:**

2. $g_1(\vec{a}_1) = 1.76 \implies A$

$g_1(\vec{a}_2) = 2.24 \implies A$

$g_1(\vec{b}_1) = 1.3 \overset{!}{\implies} A$ (Fehlklassifikation)

$g_1(\vec{b}_2) = 1.48 \overset{!}{\implies} A$ (Fehlklassifikation)

Damit sind zwei Punkte aus Klasse B falsch klassifiziert: $E_1 = \{\vec{b}_1, \vec{b}_2\}$

$$\begin{pmatrix} \vec{w}_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} -0.2 \\ -0.4 \\ 2 \end{pmatrix} - (1)\vec{b}_1 - (1)\vec{b}_2 = \begin{pmatrix} -0.2 \\ -0.4 \\ 2 \end{pmatrix} - (1)\begin{pmatrix} 0.4 \\ 1.6 \\ 1 \end{pmatrix} - (1)\begin{pmatrix} -0.4 \\ 1.5 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.1 \\ -3.5 \\ 0 \end{pmatrix}$$

$\implies g_2(x) = -0.1x_1 - 3.5x_2$

Geradengleichung für die Zeichnung: $x_2 = \frac{-x_1}{35}$

Exercise 5

- **Iteration 3 & 4 & 5:**

3. $g_2(\vec{a}_1) = -2.76 \overset{!}{\Longrightarrow} A$
   $g_2(\vec{a}_2) = 0.78 \Longrightarrow A$
   $g_2(\vec{b}_1) = -5.63 \Longrightarrow B$
   $g_2(\vec{b}_2) = -5.21 \Longrightarrow B$

   $$\begin{pmatrix} \vec{w}_3 \\ b_3 \end{pmatrix} = \begin{pmatrix} -0.5 \\ -2.7 \\ 1 \end{pmatrix}$$

   Geradengleichung für die Zeichnung: $x_2 = \frac{-5x_1+10}{27}$

4. $g_3(\vec{a}_1) = -0.96 \overset{!}{\Longrightarrow} A$
   $g_3(\vec{a}_2) = 1.94 \Longrightarrow A$
   $g_3(\vec{b}_1) = -4.47 \Longrightarrow B$
   $g_3(\vec{b}_2) = -2.85 \Longrightarrow B$

   $$\begin{pmatrix} \vec{w}_4 \\ b_4 \end{pmatrix} = \begin{pmatrix} -0.9 \\ -1.9 \\ 2 \end{pmatrix}$$

   Geradengleichung für die Zeichnung: $x_2 = \frac{-9x_1+20}{19}$

5. $g_4(\vec{a}_1) = 0.84 \Longrightarrow A$
   $g_4(\vec{a}_2) = 3.1 \Longrightarrow A$
   $g_4(\vec{b}_1) = -1.31 \Longrightarrow B$
   $g_4(\vec{b}_2) = -0.49 \Longrightarrow B$

- Stopped at
- $(w_4, b_4) = (-0.9, -1.9, 2.0)^T$

Exercise 5