Klasifikace a rozpoznávání

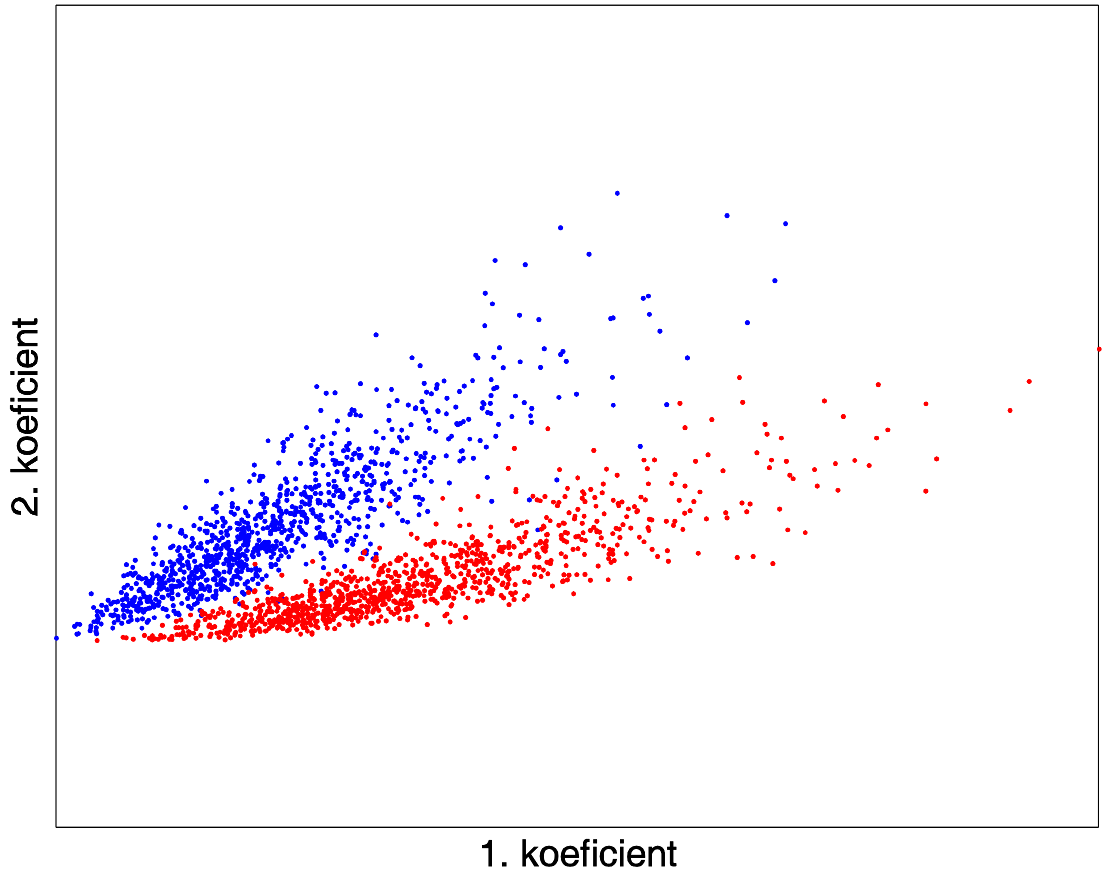
Extrakce příznaků

Extrakce příznaků - parametrizace

- Poté co jsme ze snímače obdržely data která jsou relevantní pro naši klasifikační úlohu, je potřeba je přizpůsobit potřebám rozpoznávače
- Klasifikátory mají rády parametry které jsou:
  - Gaussovského rozložení (většinou vícerozměrného)
  - Nekorelované
  - Nízkodimenzionální

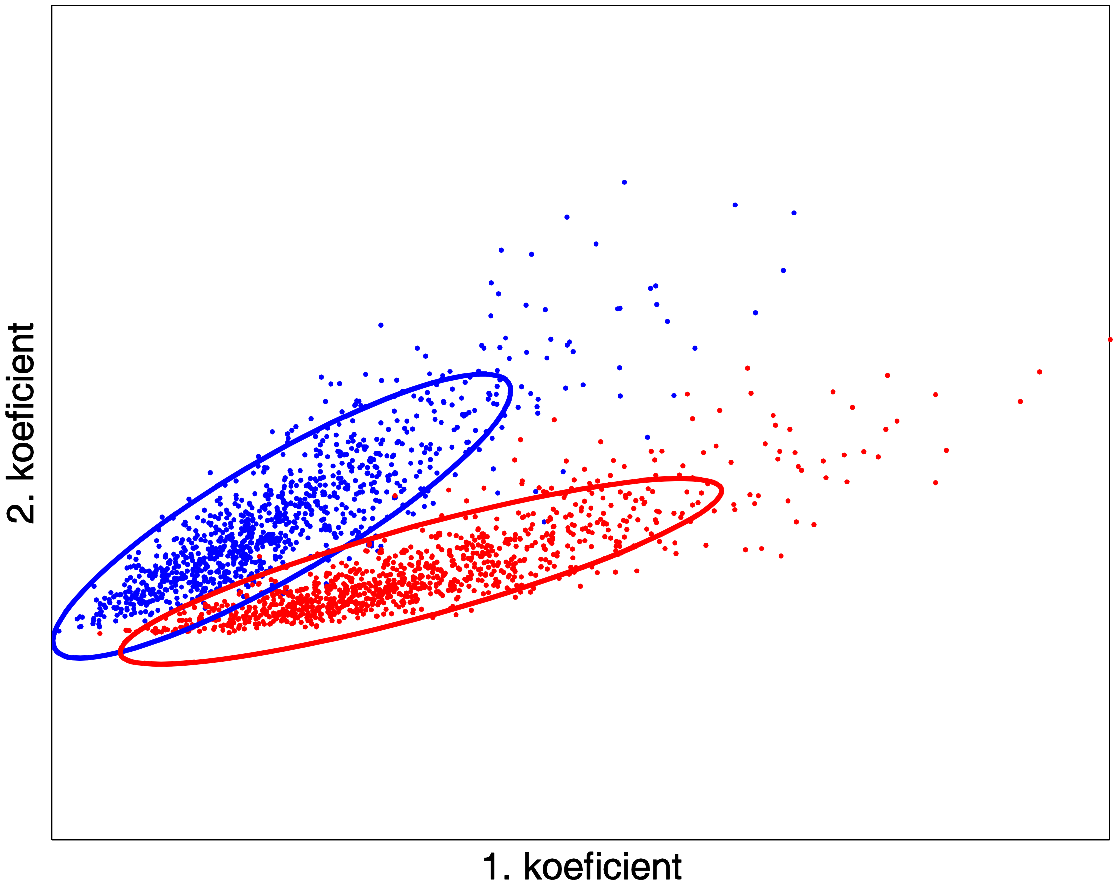
Příklad parametrizace pro 2D vstupní vektory

- Mějme vzorky (příklady) 2D rozložení pro dvě třídy.



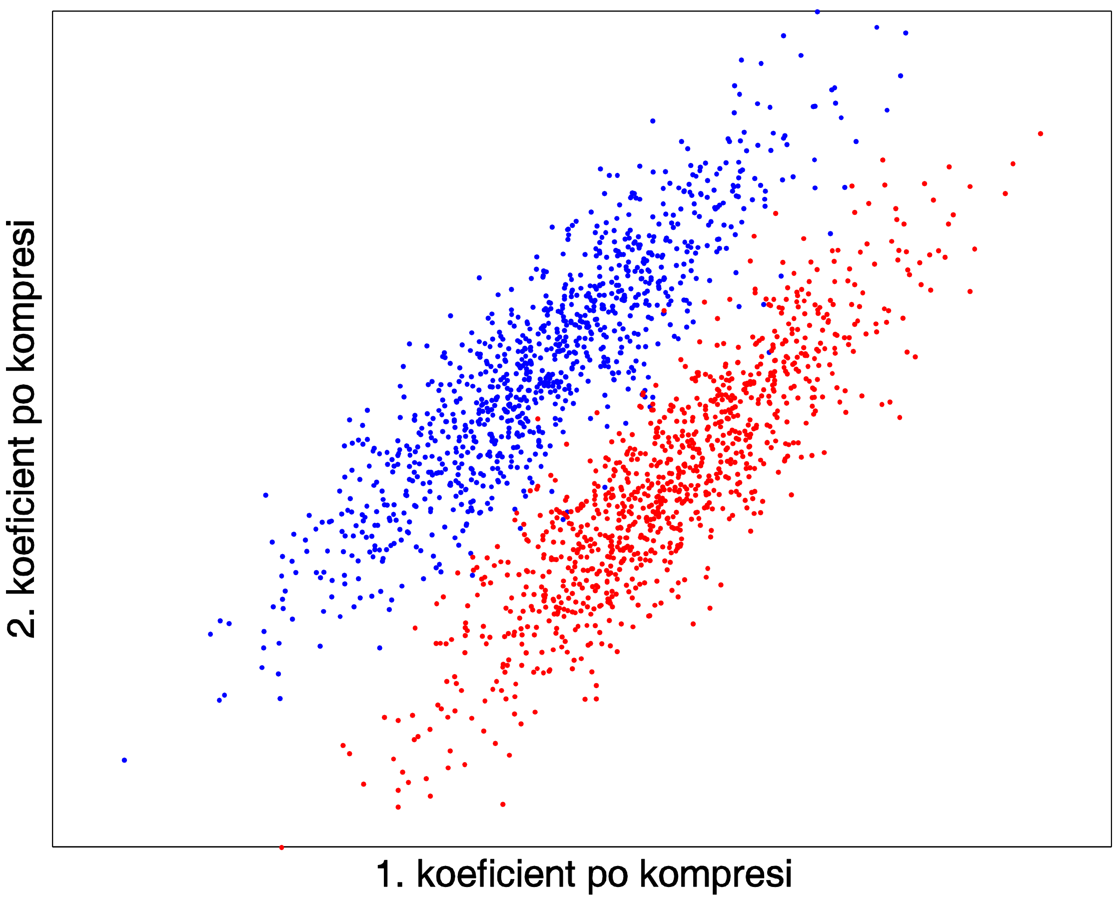
Příklad parametrizace pro 2D vstupní vektory

- Rozložení není příliš gaussovské.
- Provedeme třetí odmocninou obou koeficientů.



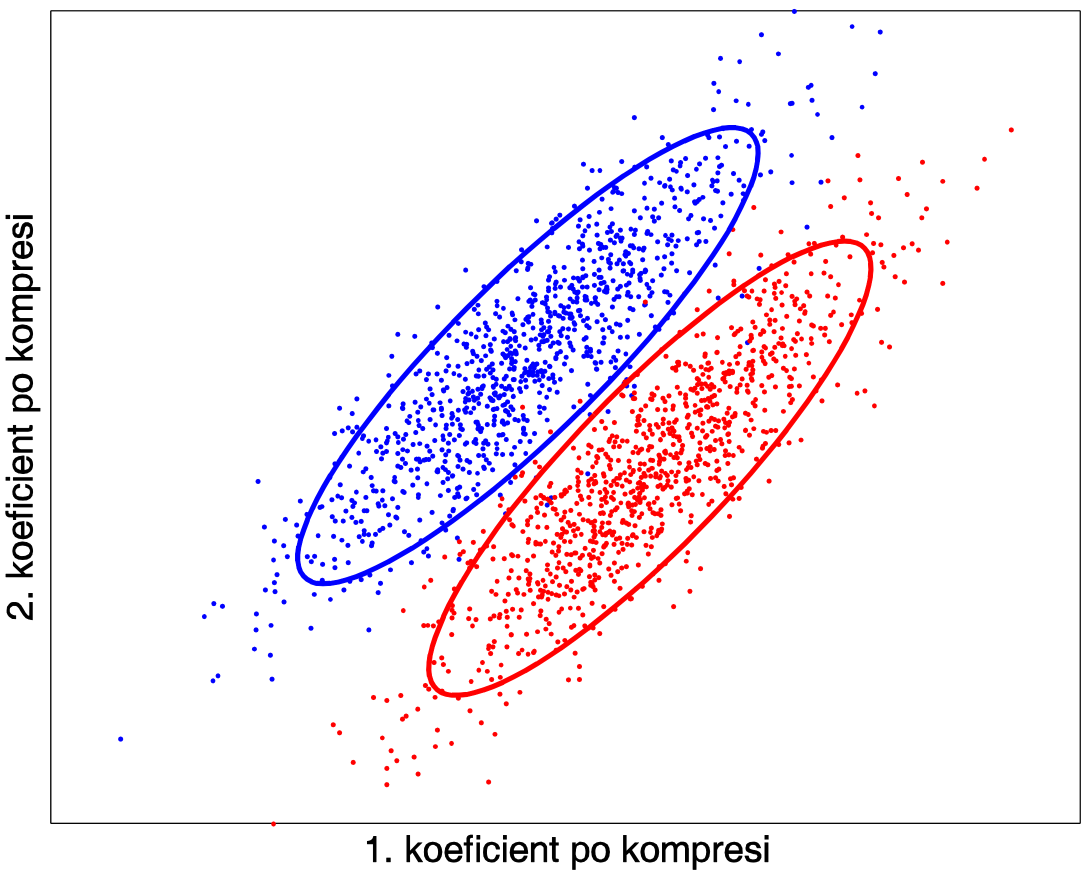
Příklad parametrizace pro 2D vstupní vektory

- Prostor se komprimuje – nelineárně deformuje...



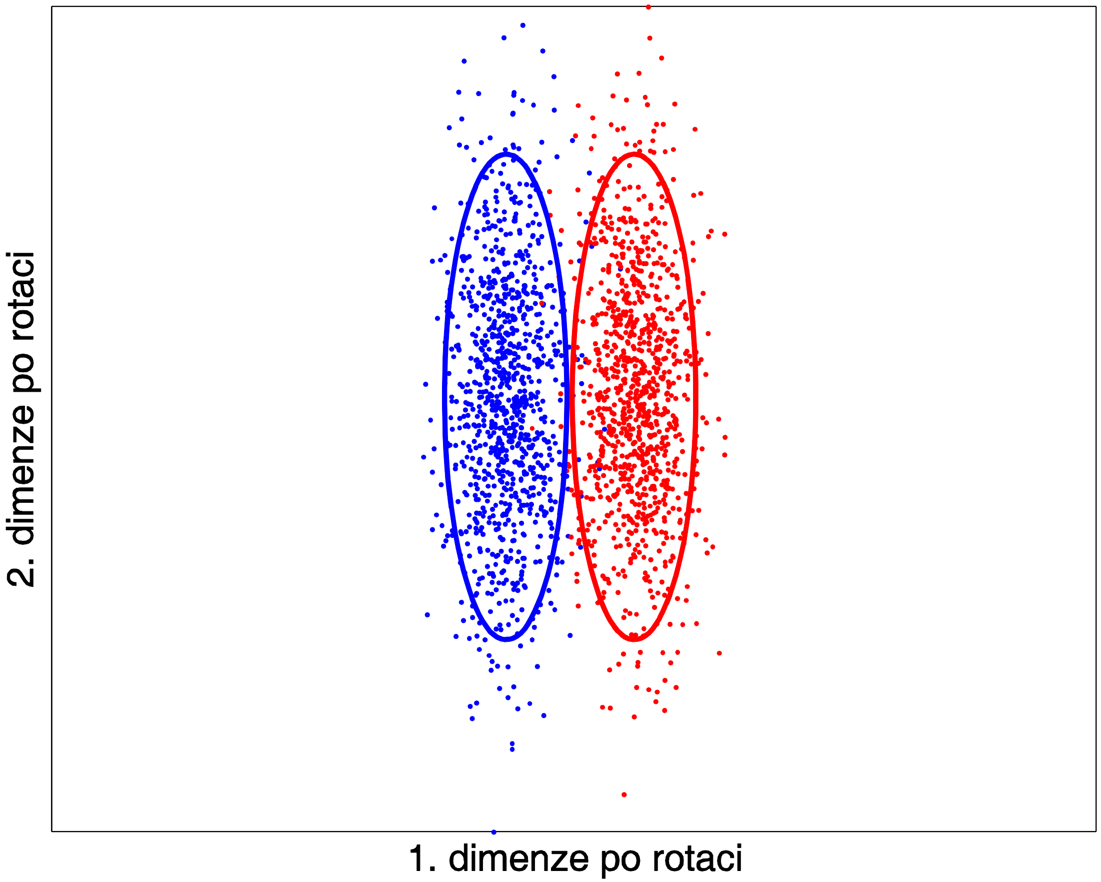
Příklad parametrizace pro 2D vstupní vektory

- ... a rozložení pro každou třídu je nyní gaussovské.
- Koeficienty jsou ale korelované.
- Je vhodné prostor otočit tak aby se koeficienty dekorelovaly.



Příklad parametrizace pro 2D vstupní vektory

- Nyní jsou koeficienty dekorelovány.
- Svislá dimenze je navíc zbytečná, protože třídy se v ní zcela překrývají.



Gaussovské rozložení (jednorozměrné)

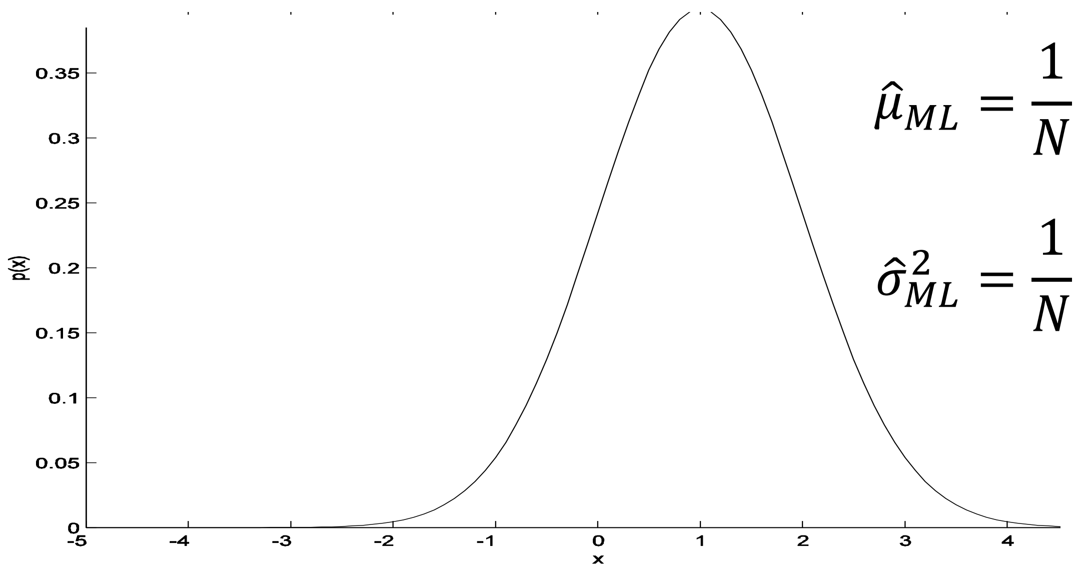
Evaluation:

$$p(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

ML odhad parametrů (Trénování):

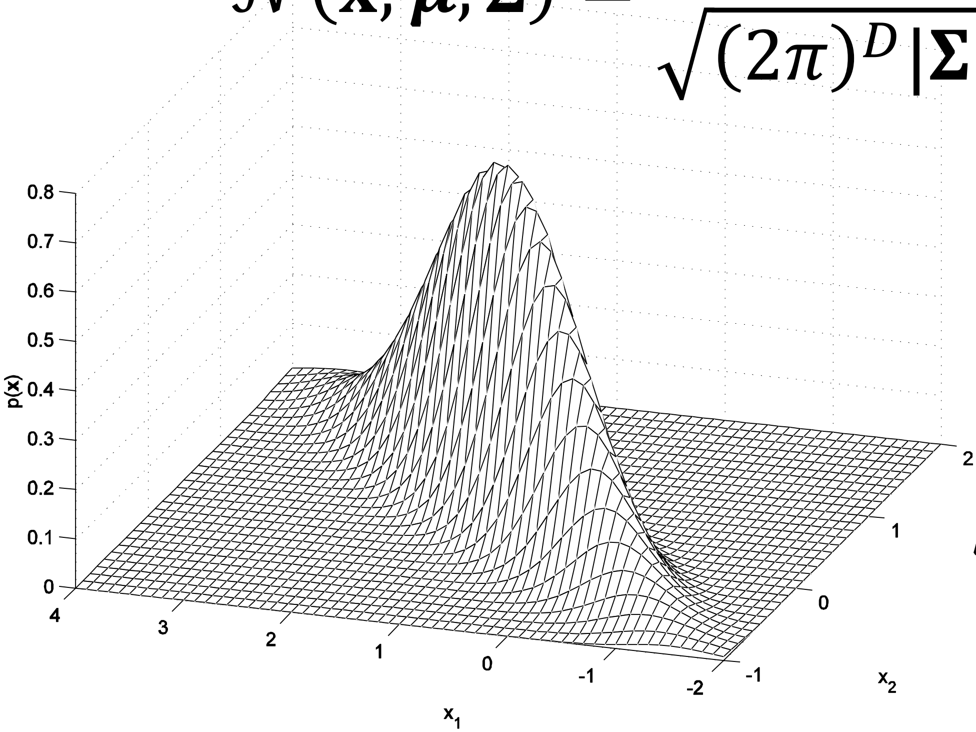
$$\hat{\mu}_{ML} = \frac{1}{N}\sum_n x_n$$

$$\hat{\sigma}^2_{ML} = \frac{1}{N}\sum_n (x_n - \hat{\mu}_{ML})^2 = \frac{1}{N}\sum_n x_n^2 - \hat{\mu}^2_{ML}$$



Gaussovské rozložení
(vícerozměrné)

$$p(x_1, \ldots, x_D) =$$
$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

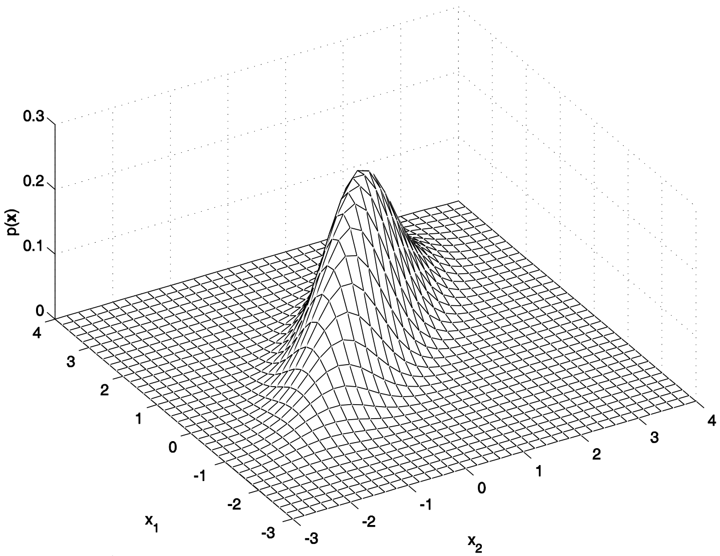


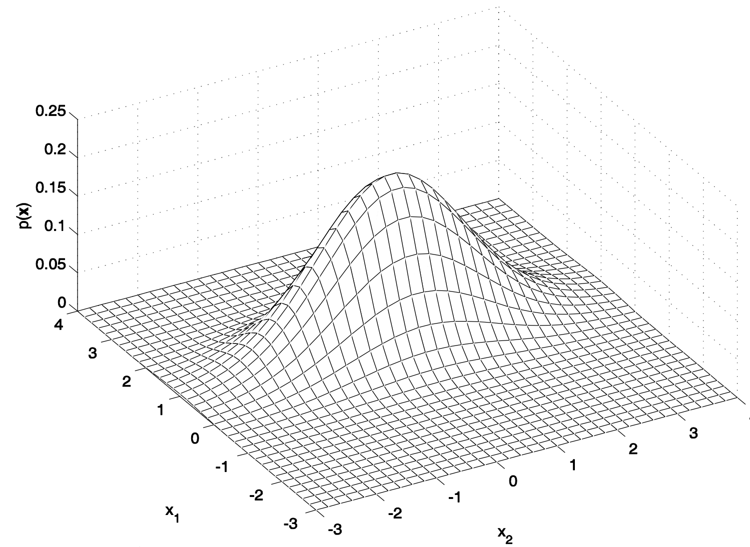
ML odhad of parametrů:

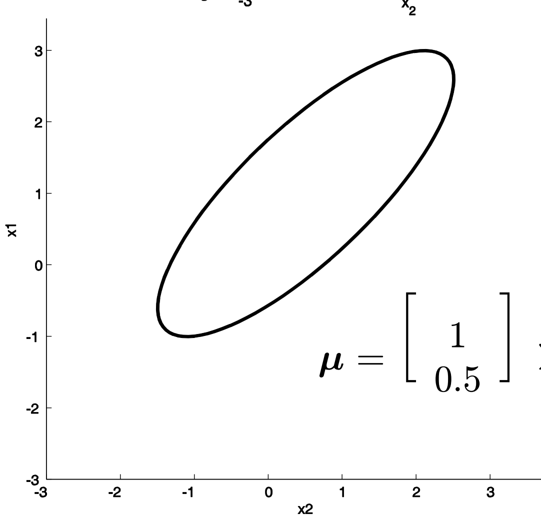
$$\boldsymbol{\mu} = \frac{1}{T}\sum_i \mathbf{x}_i$$

$$\boldsymbol{\Sigma} = \frac{1}{T}\sum_i (\mathbf{x}_\mathrm{i} - \boldsymbol{\mu})(\mathbf{x}_\mathrm{i} - \boldsymbol{\mu})^T$$

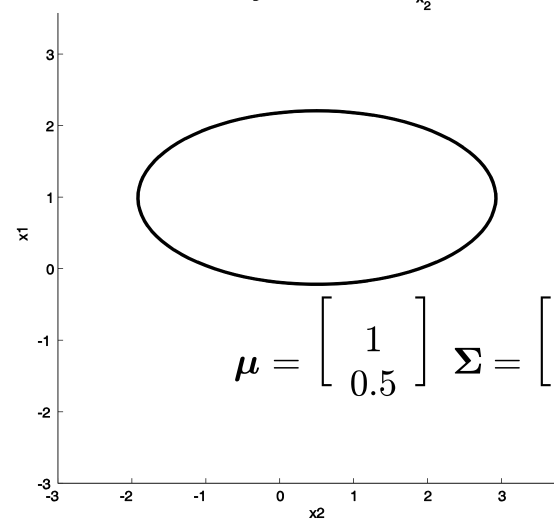
Plná a diagonální kovarianční matice







$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$



$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.4 & 0 \\ 0 & 1.5 \end{bmatrix}$$

Diagonální kovarianční matice

$$\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Pokud je $\boldsymbol{\Sigma}$ diagonální matice s koeficienty v diagonále σ_i^2 ➔

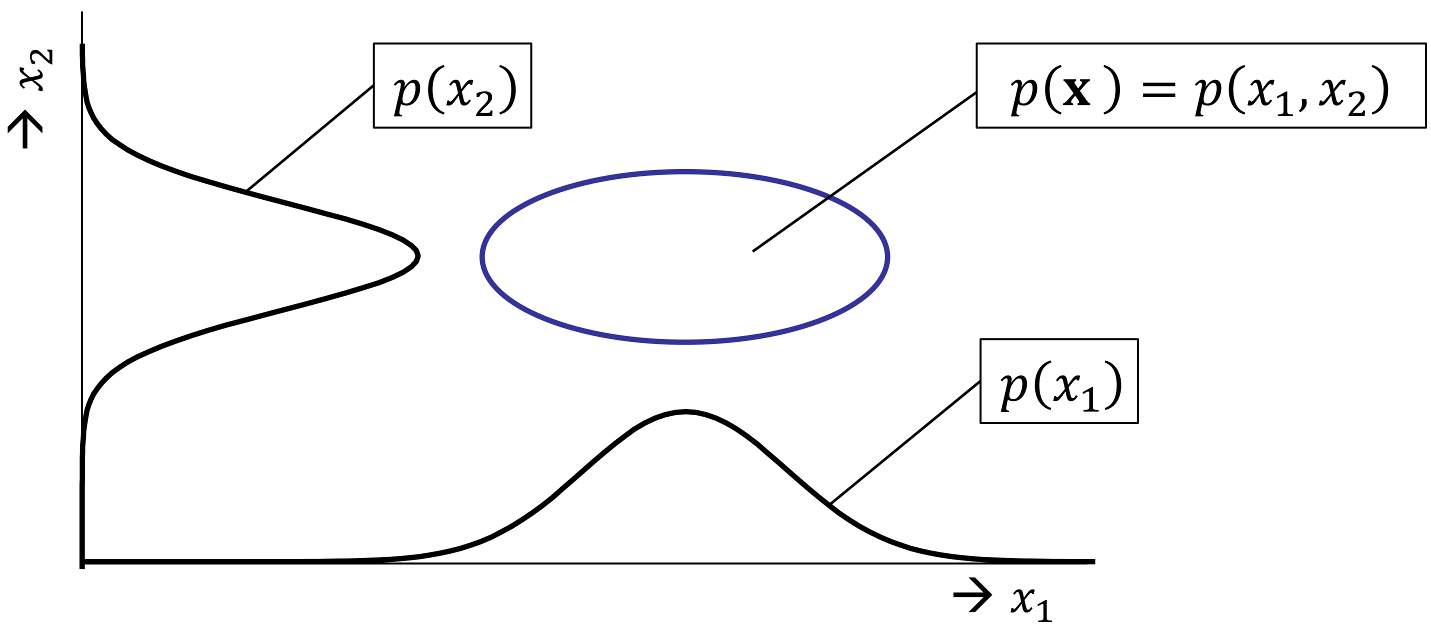
$$\begin{aligned}\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \frac{1}{\sqrt{(2\pi)^D\prod_{i=1}^{D}\sigma_i^2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{D}\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right\} \\ &= \left(\prod_{i=1}^{D}\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)\left(\prod_{i=1}^{D}\exp\left\{-\frac{1}{2}\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right\}\right) \\ &= \prod_{i=1}^{D}\mathcal{N}\left(x_i;\mu_i,\sigma_i^2\right)\end{aligned}$$

Diagonální kovarianční matice

$P(A,B) = P(A)P(B) \;\Rightarrow$ Jevy A a B jsou statisticky nezávislé

$$\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{i=1}^{D} \mathcal{N}\left(x_i;\mu_i,\sigma_i^2\right)$$

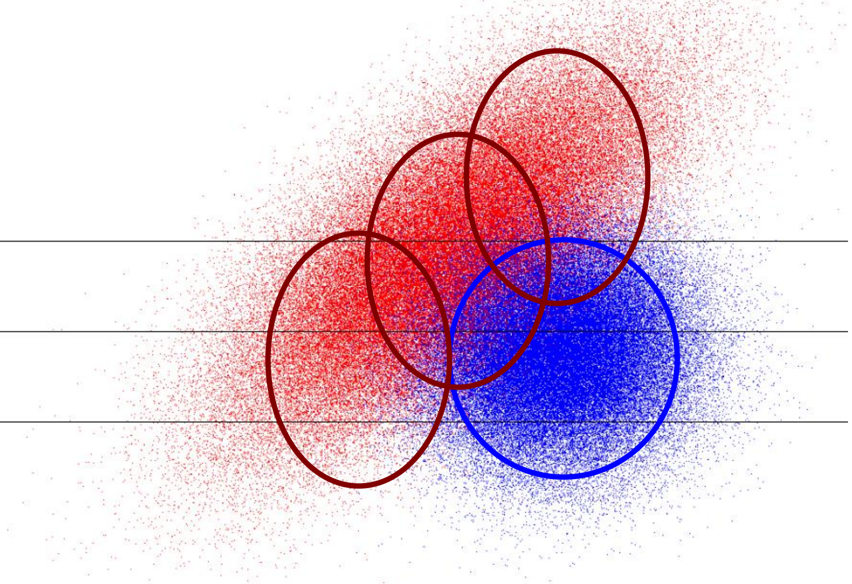
Koeficienty x_i příznakového vektoru **x** jsou statisticky nezávislé.

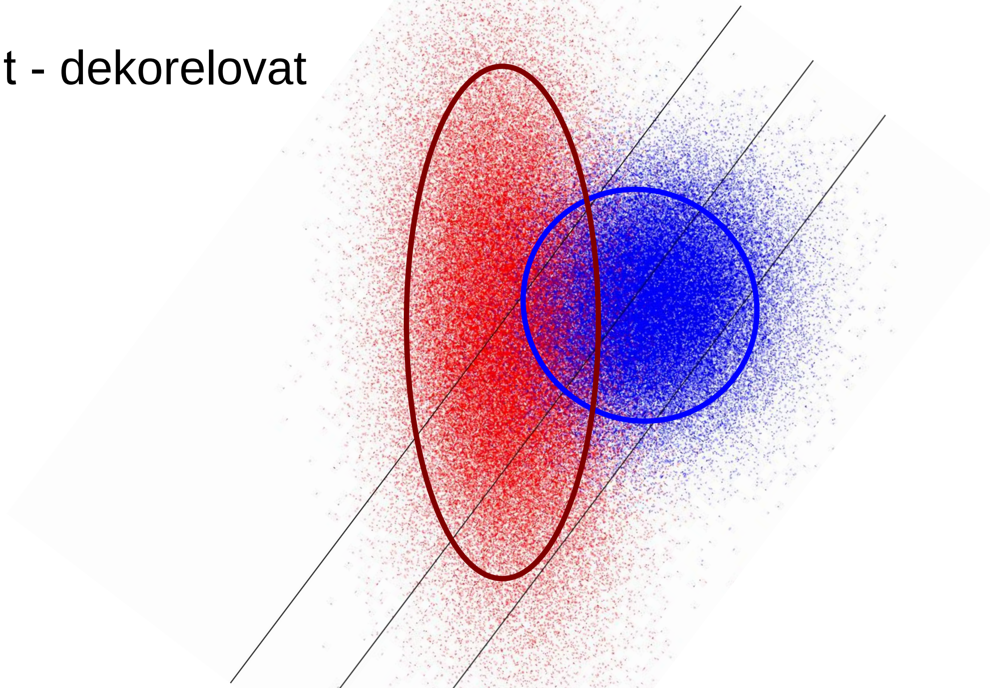


Diagonální kovarianční matice

Proč nás zajímá?

- Pomůže nám pochopit význam plné kovarianční matice v gaussovském rozložení
- Úspora parametrů při modelování dat
- Pokud jsou data korelována (viz červená třída na prvním obr.)
  - Zvláště pro vysoce dimenzionální příznaky, modelování pomocí směsi gaussovských rozložení s diagonální $\boldsymbol{\Sigma}$ může být úspornější než použití jedné gaussovky s plnou $\boldsymbol{\Sigma}$
  - Můžeme se pokusit data natočit - dekorelovat

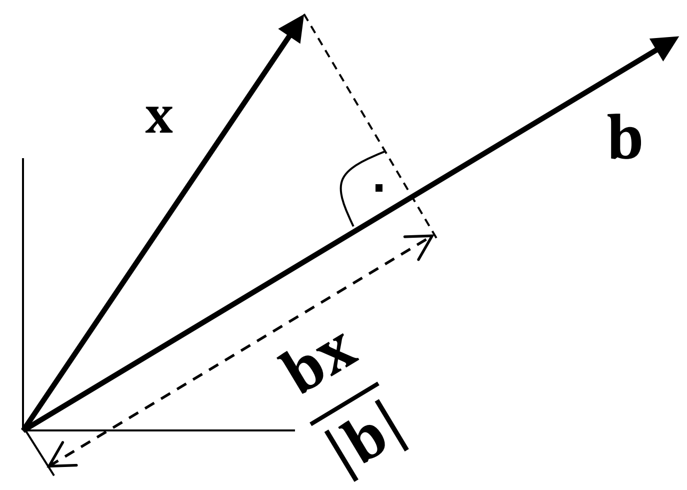




Skalární součin

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_1 & b_2 \end{bmatrix}$$

$$\mathbf{bx} = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b_1 x_1 + b_2 x_2$$



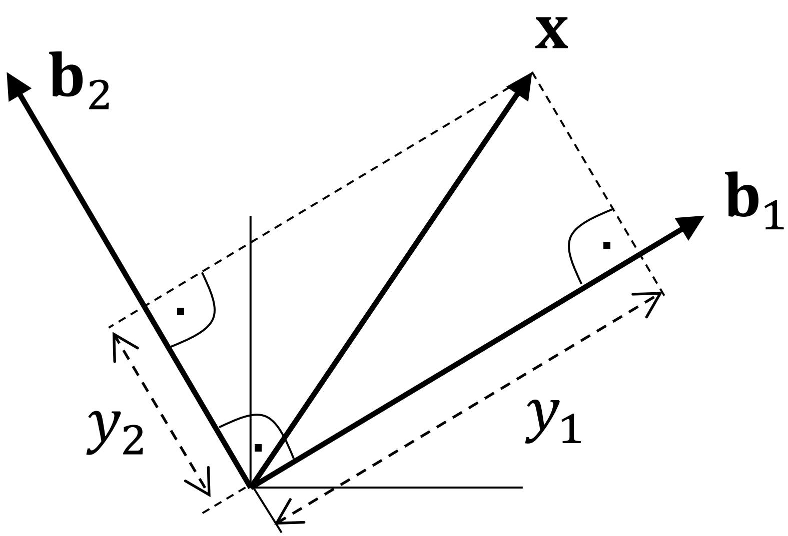
Rotace vektoru

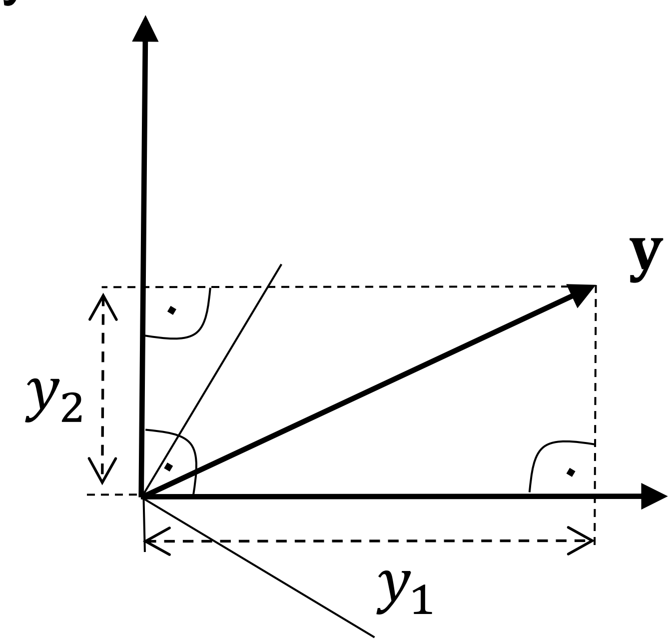
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Nechť $\mathbf{b}_1$ a $\mathbf{b}_2$ jsou ortonormální baze

- Vektory jsou na sebe kolmé
- Mají délku $|\mathbf{b}_1| = |\mathbf{b}_2| = 1$

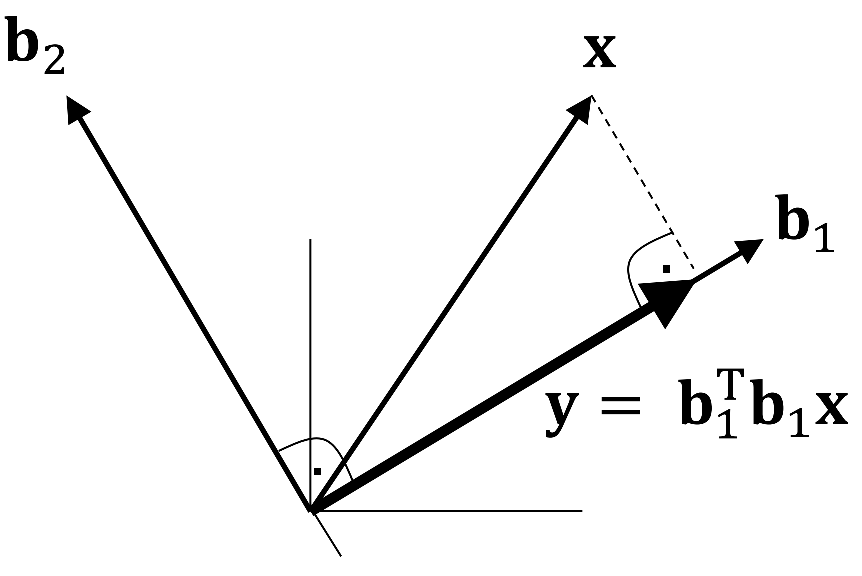
Potom $\mathbf{y} = \mathbf{B}\mathbf{x}$ je otočený vektor $\mathbf{x}$, kde $\mathbf{b}_1$ a $\mathbf{b}_2$ ukazují v původním prostoru směry nových os





Projekce vektoru

- Nechť $\mathbf{B}$ je matice ortonormálních bází a $\overline{\mathbf{B}}$ matice tvořena pouze několika řádky (bázemi) matice $\mathbf{B}$.
- Potom $\mathbf{y} = \overline{\mathbf{B}}^T \overline{\mathbf{B}} \mathbf{x}$ je projekce vektoru $\mathbf{x}$ do bází $\overline{\mathbf{B}}$.



Vlastní čísla a vektory

λ je vlastní číslo a $\mathbf{e}$ je odpovídající vlastni vektor čtvercové matice $\mathbf{\Sigma}$, pokud platí:

$$\mathbf{\Sigma e} = \mathbf{e}\lambda$$

$\mathrm{D} \times \mathrm{D}$ matice má (nanejvýš) D různých vlastních čísel. Nechť je $\mathbf{\Lambda}$ diagonální matice všech vlastních čísel a matice $\mathbf{E}$ obsahuje ve sloupcích odpovídající vlastní vektory.

$$\mathbf{\Sigma E} = \mathbf{E\Lambda}$$

Nás bude zajímat speciální případ kdy matice $\mathbf{\Sigma}$ je symetrická. Potom budou sloupce matice $\mathbf{E}$ tvořit ortonormální báze. Pro takovou matici potom platí: $\mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{E}^{-1}\mathbf{E} = \mathbf{I}$ kde $\mathbf{I}$ je jednotková matice. Tedy platí následující rozklady matic:

$$\mathbf{E}^{\mathrm{T}}\mathbf{\Sigma E} = \mathbf{\Lambda} \qquad\qquad \mathbf{\Sigma} = \mathbf{E\Lambda E}^{\mathrm{T}}$$

$\boldsymbol{\mu}$ transformovaných dat

- Jak se změní odhady střední hodnoty a kovarianční matice pokud původní data transformujeme: $\mathbf{y} = \mathbf{A}\mathbf{x}$

$$
\begin{aligned}
\boldsymbol{\mu}_y &= \frac{1}{N}\sum_{n=1}^{\mathrm{T}} \mathbf{A}\mathbf{x}_{\mathrm{n}} \\
&= \mathbf{A}\frac{1}{N}\sum_{n=1}^{\mathrm{T}} \mathbf{x}_{\mathrm{n}} \\
&= \mathbf{A}\boldsymbol{\mu}_x
\end{aligned}
$$

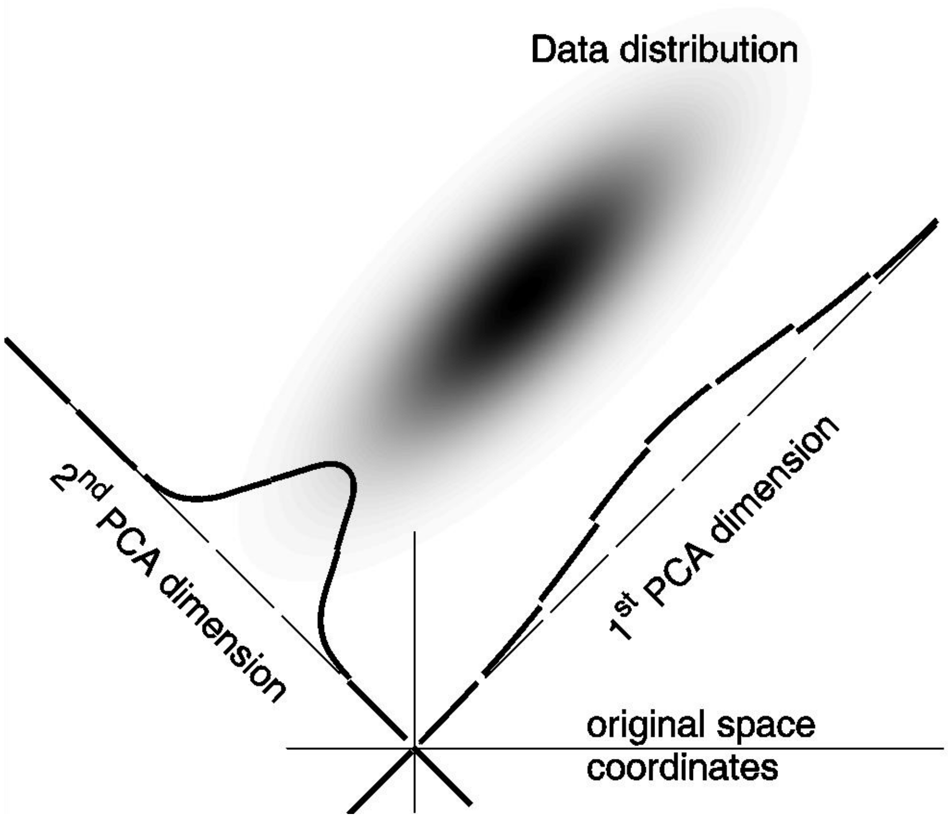
$\mathbf{\Sigma}$ transformovaných dat

$$
\begin{aligned}
\mathbf{\Sigma}_y &= \frac{1}{N}\sum_{n=1}^{\mathrm{T}}\left(\mathbf{A}\mathbf{x}_{\mathrm{n}}-\boldsymbol{\mu}_y\right)\left(\mathbf{A}\mathbf{x}_{\mathrm{n}}-\boldsymbol{\mu}_y\right)^{\mathrm{T}} \\
&= \frac{1}{N}\sum_{n=1}^{\mathrm{T}}(\mathbf{A}\mathbf{x}_{\mathrm{n}}-\mathbf{A}\boldsymbol{\mu}_x)(\mathbf{A}\mathbf{x}_{\mathrm{n}}-\mathbf{A}\boldsymbol{\mu}_x)^{\mathrm{T}} \\
&= \mathbf{A}\left\{\frac{1}{N}\sum_{n=1}^{\mathrm{T}}(\mathbf{x}_{\mathrm{n}}-\boldsymbol{\mu}_x)(\mathbf{x}_{\mathrm{n}}-\boldsymbol{\mu}_x)^{\mathrm{T}}\right\}\mathbf{A}^{\mathrm{T}} \\
&= \mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^{\mathrm{T}}
\end{aligned}
$$

Co se stane když jako **A** použijeme transponovanou matici vlastních vektoru kovarianční matice $\mathbf{\Sigma}_x$? (Proč transponovanou? Protože vlastní vektory máme ve sloupcích a ne v řádcích). Jaký význam mají vlastní čísla?

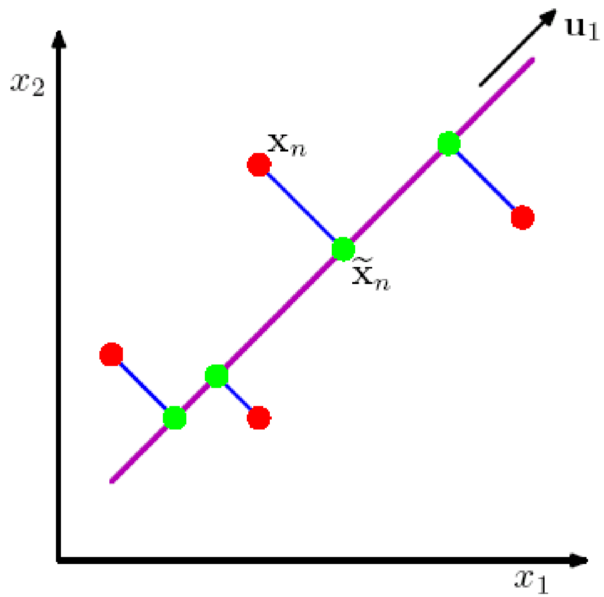
Analýza hlavních komponent

(Principal Component Analysis - PCA)



Analýza hlavních komponent

- Umožňuje:
  - Dekorelaci – vlastní vektory kovarianční matice definuji souřadný systém ve kterých jsou data dekorelována – mají diagonální kovarianční matici
  - Redukci dimenzí – promítnutí dat do pouze několika vlastních vektorů odpovídajících největším vlastním číslům (směry s nevětší variancí) umožní optimální rekonstrukci dat s nejmenší kvadratickou chybou (mean square error - MSE)
  - Redukce dimenzí provádíme pokud věříme, že v některých směrech není užitečná informace ale pouze (gaussovský) šum s nízkou variabilitou.



Interpretace **Σ** v gaussovském rozložení

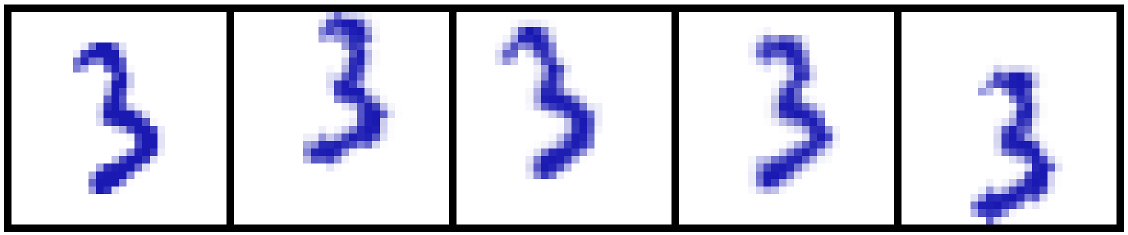
$$\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

$$= \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{E}\boldsymbol{\Lambda}^{-1}\mathbf{E}^{\mathrm{T}}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

$$= \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Lambda}|}}\exp\left\{-\frac{1}{2}\left(\mathbf{E}^{\mathrm{T}}\mathbf{x}-\mathbf{E}^{\mathrm{T}}\boldsymbol{\mu}\right)^T\boldsymbol{\Lambda}^{-1}\left(\mathbf{E}^{\mathrm{T}}\mathbf{x}-\mathbf{E}^{\mathrm{T}}\boldsymbol{\mu}\right)\right\}$$

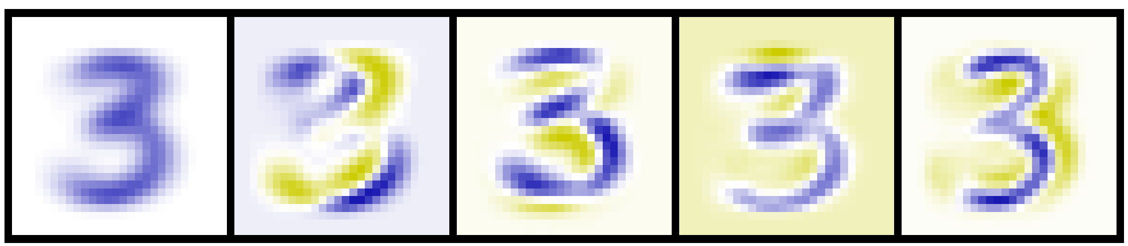
PCA - Příklad

- Obrázky 100x100 pixelů – 10000 dimensionální vektory



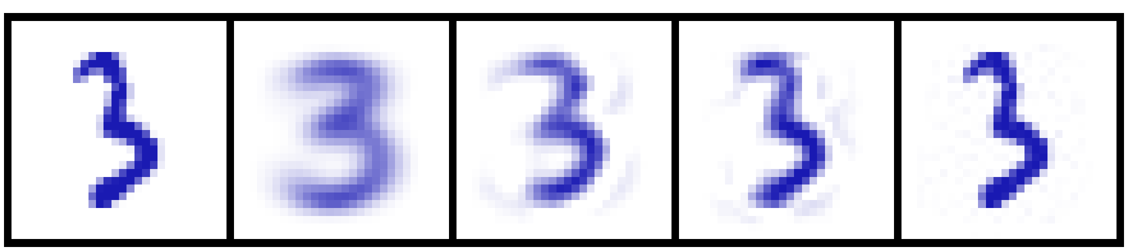
- Střední hodnota, vlastní čísla a vlastní vektory

μ $\lambda_1=3.4\cdot10^5$ $\lambda_2=2.8\cdot10^5$ $\lambda_3=2.4\cdot10^5$ $\lambda_3=1.6\cdot10^5$



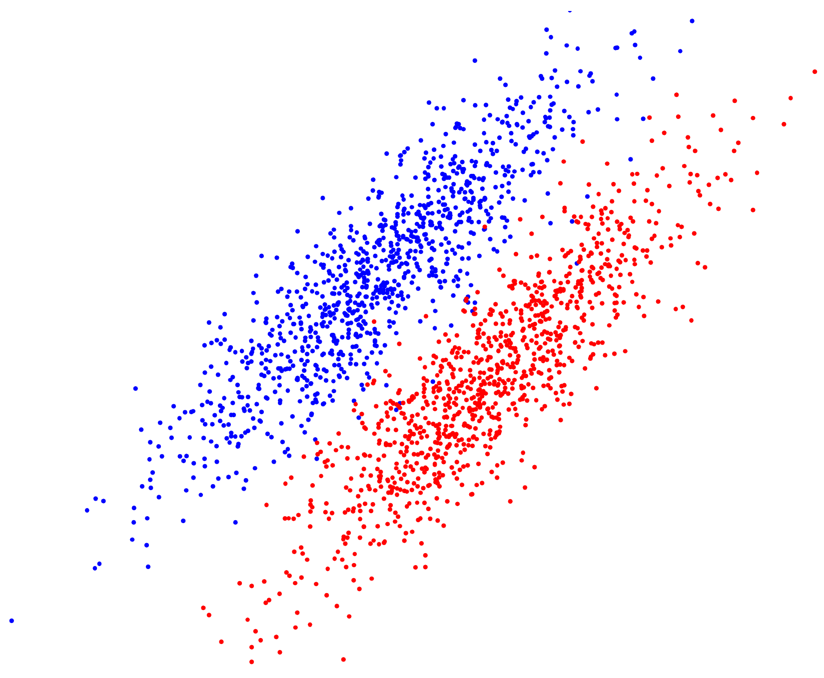
- Střední hodnota, vlastní čísla a vlastní vektory

Originál M = 1 M=10 M=50 M=250



PCA - Příklad

- Jakou dimenzi si PCA vybere na tomto příkladě?
- Bude to výhodné pro klasifikaci tříd?



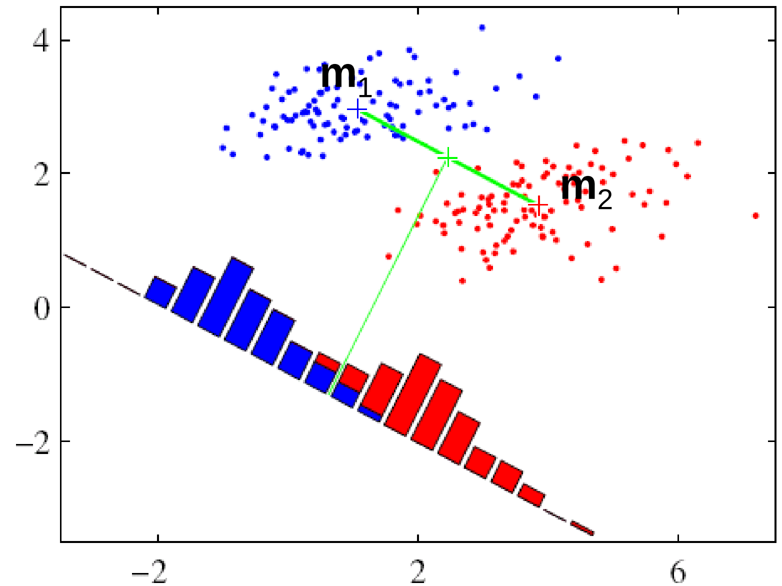
Lineární diskriminační analýza

- Opět se pokusíme promítnout data pouze do určitého směru: $y = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

- Tentokrát ale budeme chtít aby v tomto směru byly separovány třídy.
- Intuitivně by nás mohlo napadnout vybrat směr ve kterém jsou nejlépe odděleny průměty středních hodnot tříd $\mathbf{m}_1$ a $\mathbf{m}_2$. Hledáme tedy $\mathbf{w}$, které maximalizuje:

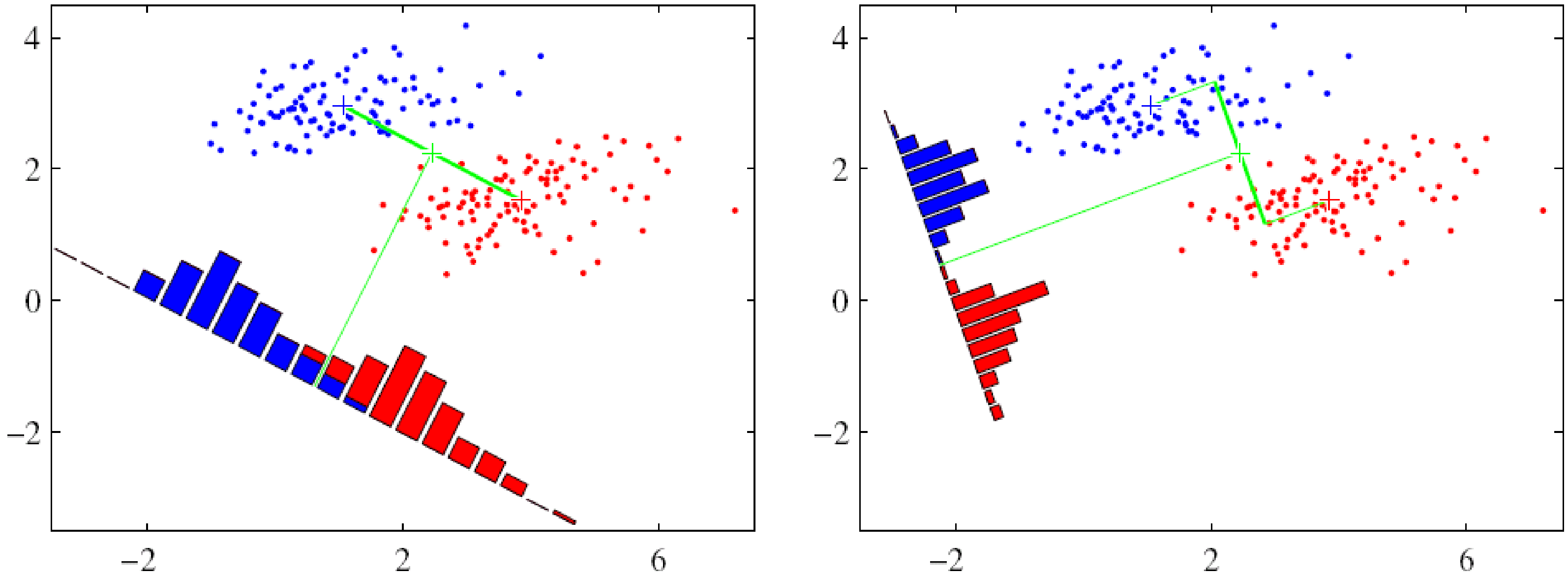
$$m_2 - m_1 = \mathbf{w}^{\mathrm{T}}(\mathbf{m}_2 - \mathbf{m}_1)$$

$$m_k = \mathbf{w}^{\mathrm{T}}\mathbf{m}_k$$



Lineární diskriminační analýza

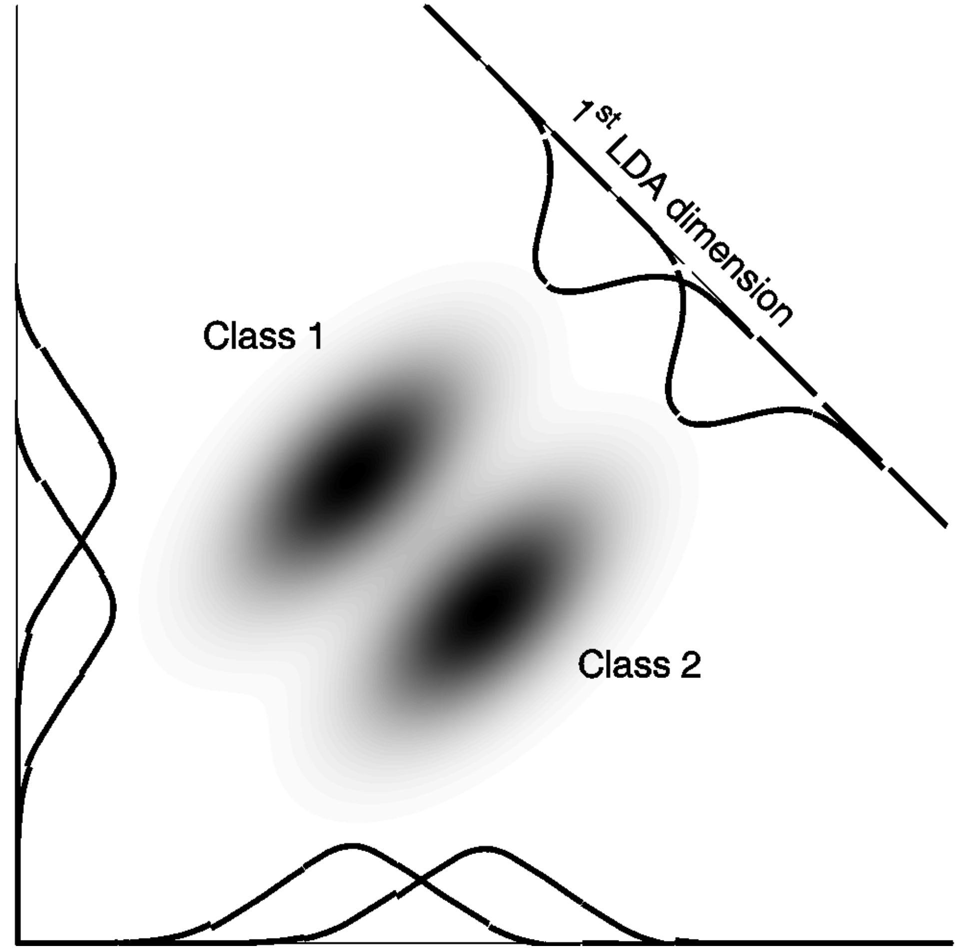
- Lze však najít i lepší směr:



- Snažíme se data promítnout do takového směru, kde
  - Maximalizujeme (čtverec) vzdálenosti mezi středními hodnotami tříd
    - Nebo-li varianci mezi třídami
  - Minimalizujeme průměrnou varianci tříd
- Maximalizujeme tedy

$$J(\mathbf{w}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2} \qquad s_k^2 = \sum_{n \in \mathcal{C}_k} (y_n - m_k)^2$$

Lineární diskriminační analýza



Lineární diskriminační analýza

- LDA dimenze dány vlastními vektory matice $\boldsymbol{\Sigma}_{\mathrm{wc}}^{-1}\boldsymbol{\Sigma}_{\mathrm{ac}}$
- Σ_{ac} – kovarianční matice spočítaná se středních hodnot tříd
- Σ_{wc} – průměrná kovarianční matice tříd
- Lze zobecnit pro více tříd – vlastní vektory s největšími vlastními čísly odpovídají směrům ve kterých jsou třídy nejlépe separovány
- Pro J tříd bude pouze J-1 vlastních čísel nenulových
- Pokud mají všechny třídy gaussovské rozložení se stejnou kovarianční maticí, LDA transformace transformuje prostor tak, že mohou byt třídy optimálně modelovány gaussovským rozložení s diagonální kovarianční maticí

$$\hat{\boldsymbol{\Sigma}}_{wc} = \frac{1}{N}\sum_{j=1}^{J} N_j \hat{\boldsymbol{\Sigma}}^{(j)} \qquad \hat{\boldsymbol{\Sigma}}^{(j)} = \frac{1}{N_j}\sum_{i=1}^{N_j}(\mathbf{x}_i^{(j)} - \hat{\boldsymbol{\mu}}^{(j)})(\mathbf{x}_i^{(j)} - \hat{\boldsymbol{\mu}}^{(j)})^T$$

$$\hat{\boldsymbol{\Sigma}}_{ac} = \frac{1}{N}\sum_{j=1}^{J} N_j (\hat{\boldsymbol{\mu}}^{(j)} - \hat{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}}^{(j)} - \hat{\boldsymbol{\mu}})^T \qquad \hat{\boldsymbol{\mu}}^{(j)} = \frac{1}{N_j}\sum_{i=1}^{N_j}\mathbf{x}_i^{(j)}$$

Lineární diskriminační analýza

- Vlastních vektor $\mathbf{w}_i$ odpovídá určitému směru v prostoru vstupních dat
- Odpovídající vlastní číslo λ_i udává poměr mezi variancí
  - středních hodnot tříd
  - průměrnou variancí v rámci tříd

  ve směru $\mathbf{w}_i$

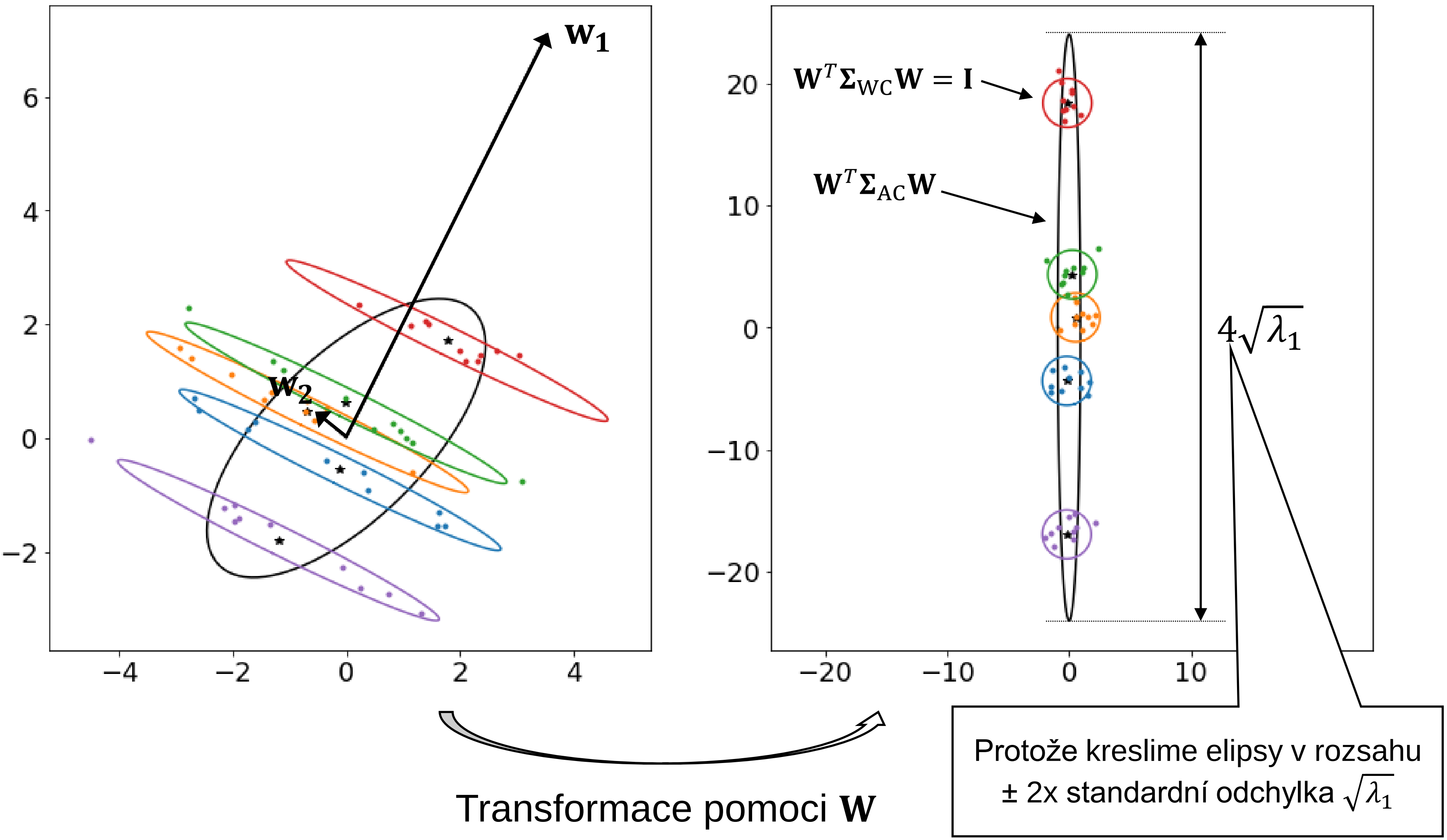
$$\lambda_i = \frac{\mathbf{w}_i^T \mathbf{\Sigma}_{\mathrm{ac}} \mathbf{w}_i}{\mathbf{w}_i^T \mathbf{\Sigma}_{\mathrm{wc}} \mathbf{w}_i}$$

- LDA najde první směr $\mathbf{w}_1$ kde je tento poměr λ_1 největší, druhý směr $\mathbf{w}_2$ kde je tento poměr λ_2 druhý největší, atd.
- Místo hledáni vlastních čísel a vektorů matice $\mathbf{\Sigma}_{\mathrm{wc}}^{-1} \mathbf{\Sigma}_{\mathrm{ac}}$ lze řešit jako *zobecněný problém vlastních čísel* (Generalized eigen value problem) pro které existují standardni implementace (*scipy.linalg.eigh*)

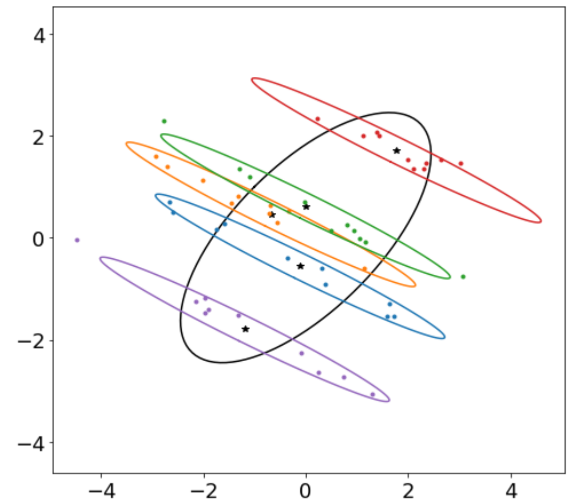
$$\mathbf{\Sigma}_{\mathrm{ac}} \mathbf{w}_i = \lambda_i \mathbf{\Sigma}_{\mathrm{wc}} \mathbf{w}_i$$

$$\mathbf{\Sigma}_{\mathrm{ac}} \mathbf{W} = \mathbf{\Sigma}_{\mathrm{wc}} \mathbf{W} \mathbf{\Lambda}$$

Lineární diskriminační analýza

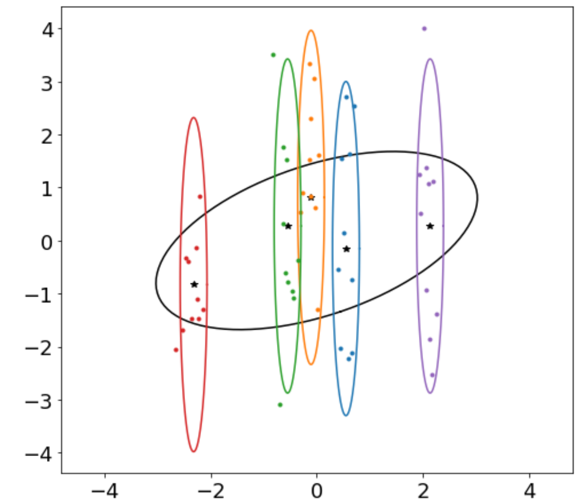


Lineární diskriminační analýza



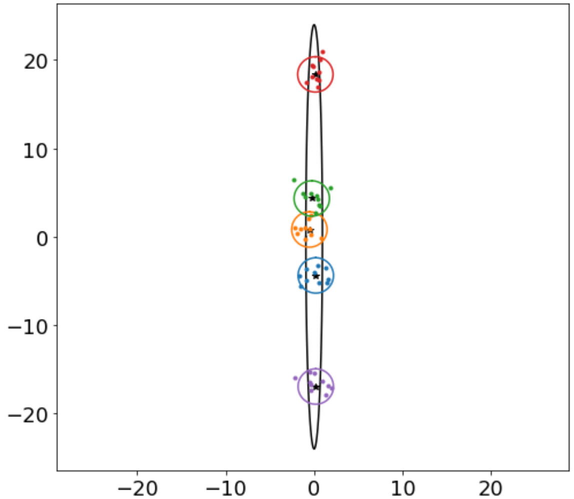
PCA pro $\mathbf{\Sigma}_{\mathrm{wc}}$

trasformace $\mathbf{E}_{\mathrm{wc}}$ jsou vlastní vektory $\mathbf{\Sigma}_{\mathrm{wc}}$



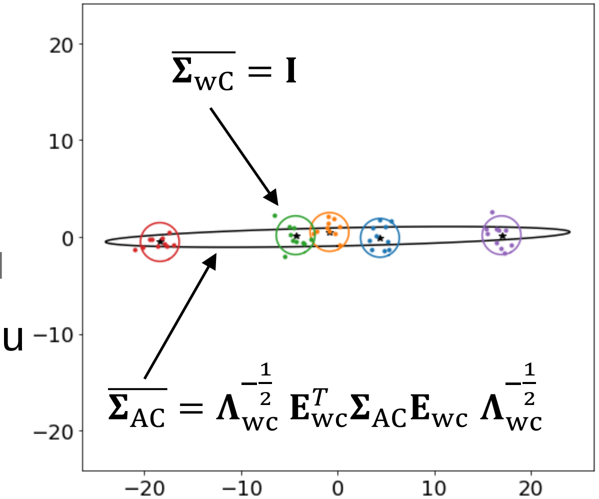
$$\mathbf{W} = \mathbf{E}_{\mathrm{AC}}^{T} \mathbf{\Lambda}_{\mathrm{wc}}^{-\frac{1}{2}} \mathbf{E}_{\mathrm{wc}}^{T}$$

trasformace $\mathbf{\Lambda}_{\mathrm{wc}}^{-\frac{1}{2}}$ jsou odmocniny inverzních vlastních čísel $\mathbf{\Sigma}_{\mathrm{wc}}$



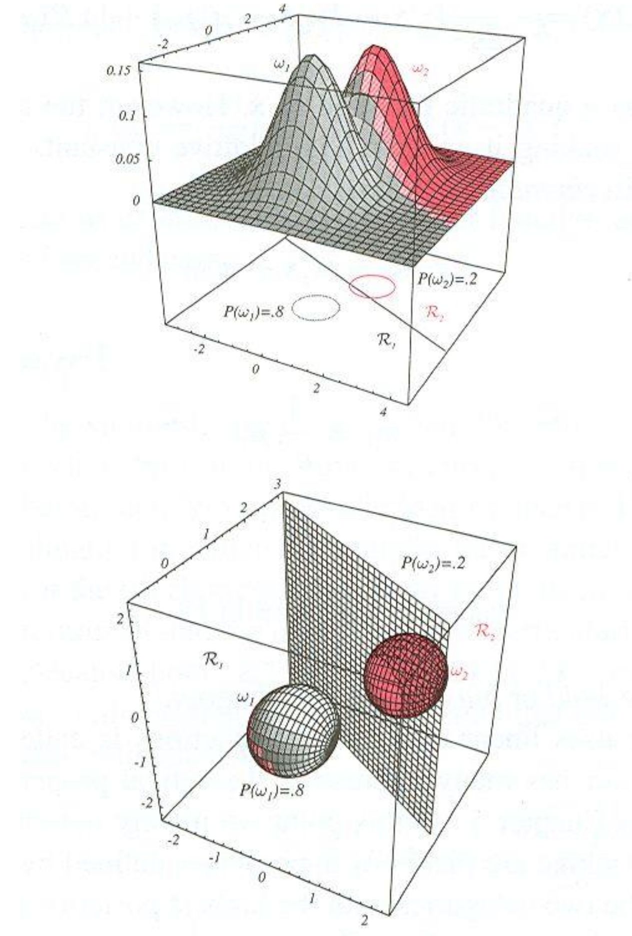
PCA pro $\overline{\mathbf{\Sigma}_{\mathrm{AC}}}$

trasformace $\mathbf{E}_{AC}$ jsou vlastní vektory $\overline{\mathbf{\Sigma}_{\mathrm{AC}}}$

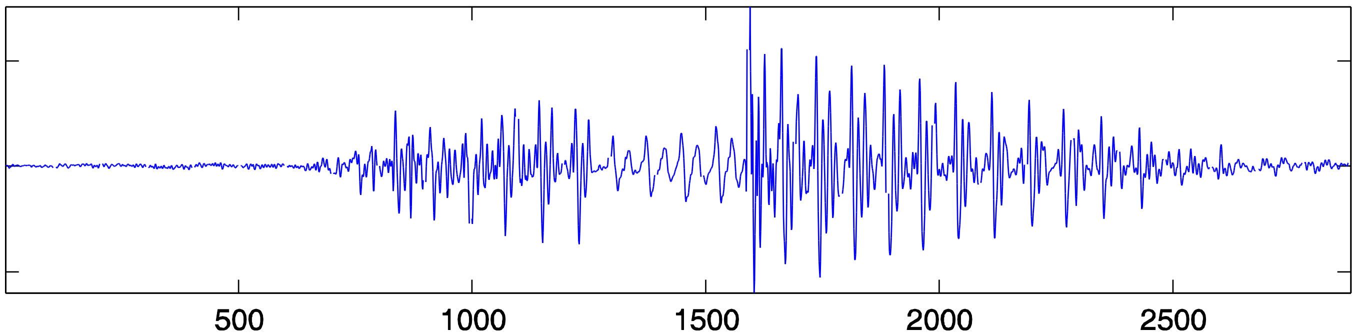


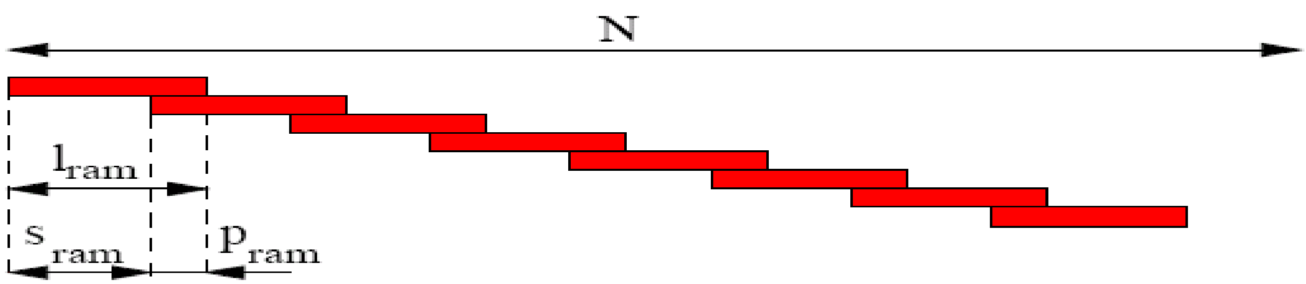
LDA a lineární klasifikátor

Dvě třídy s gaussovským rozložením se stejnou kovarianční matici jsou opravdu optimálně oddělitelné lineárním klasifikátorem (přímkou, rovinou, hyper-rovinou)

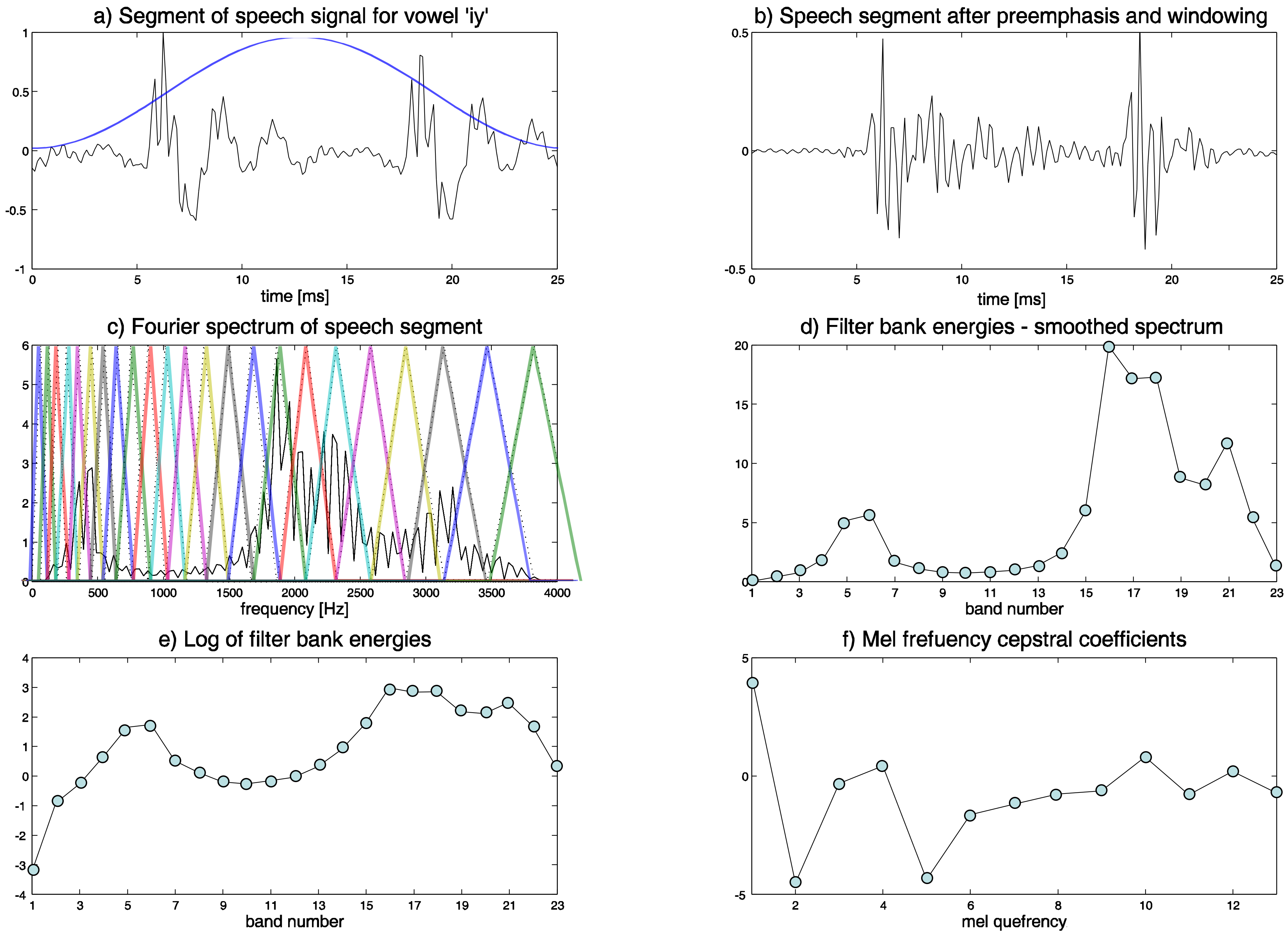


Extrakce příznaku pro řeč - MFCC (Mel frequency cepstral coefficients)

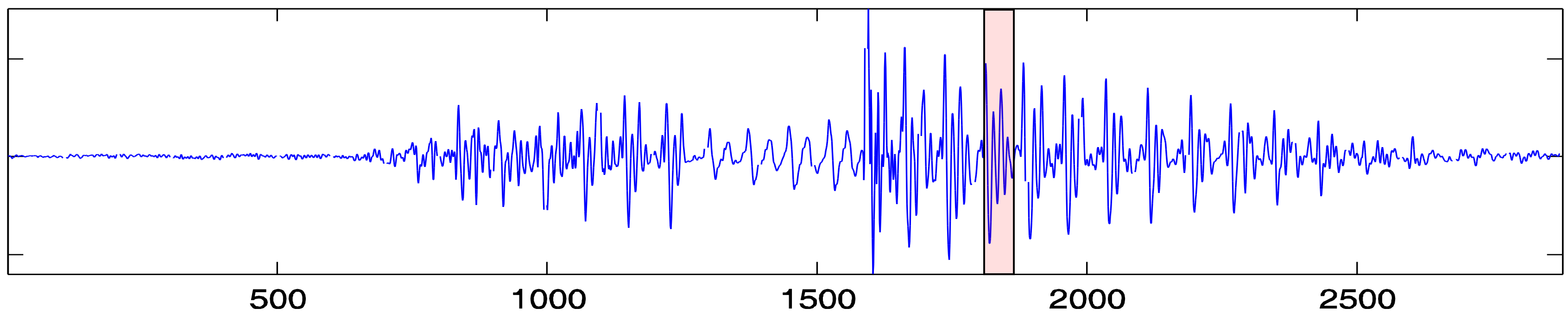




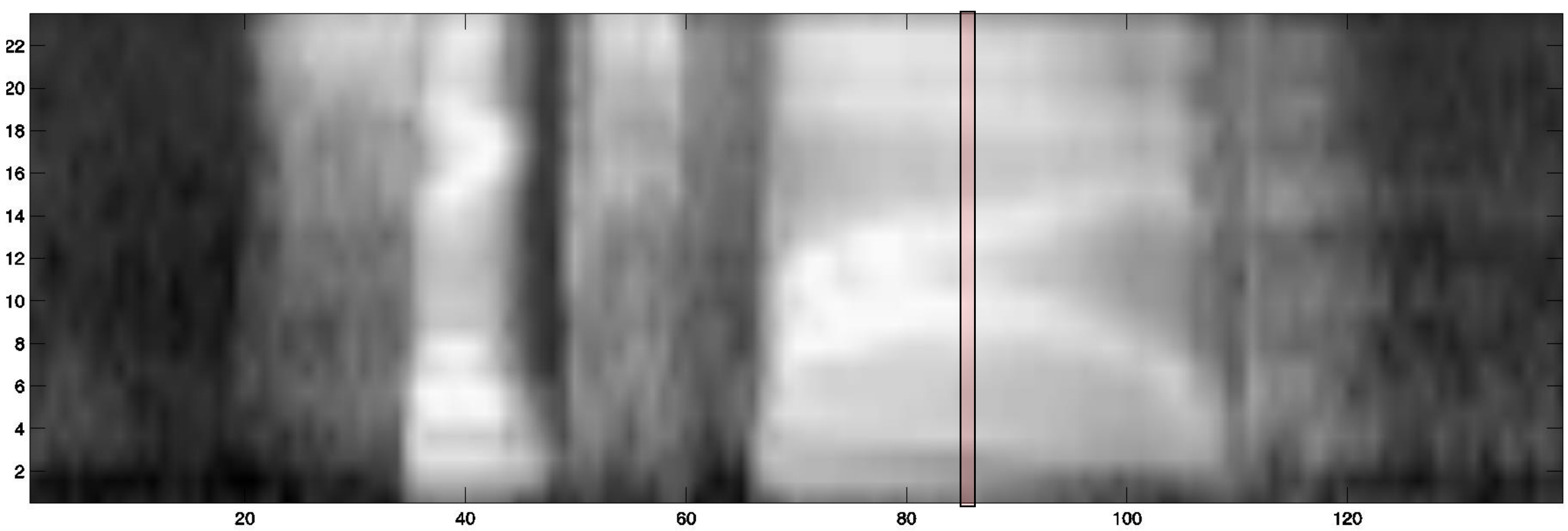
- Nejprve řečový signál rozdělíme do asi 20ms překrývajících se segmentů

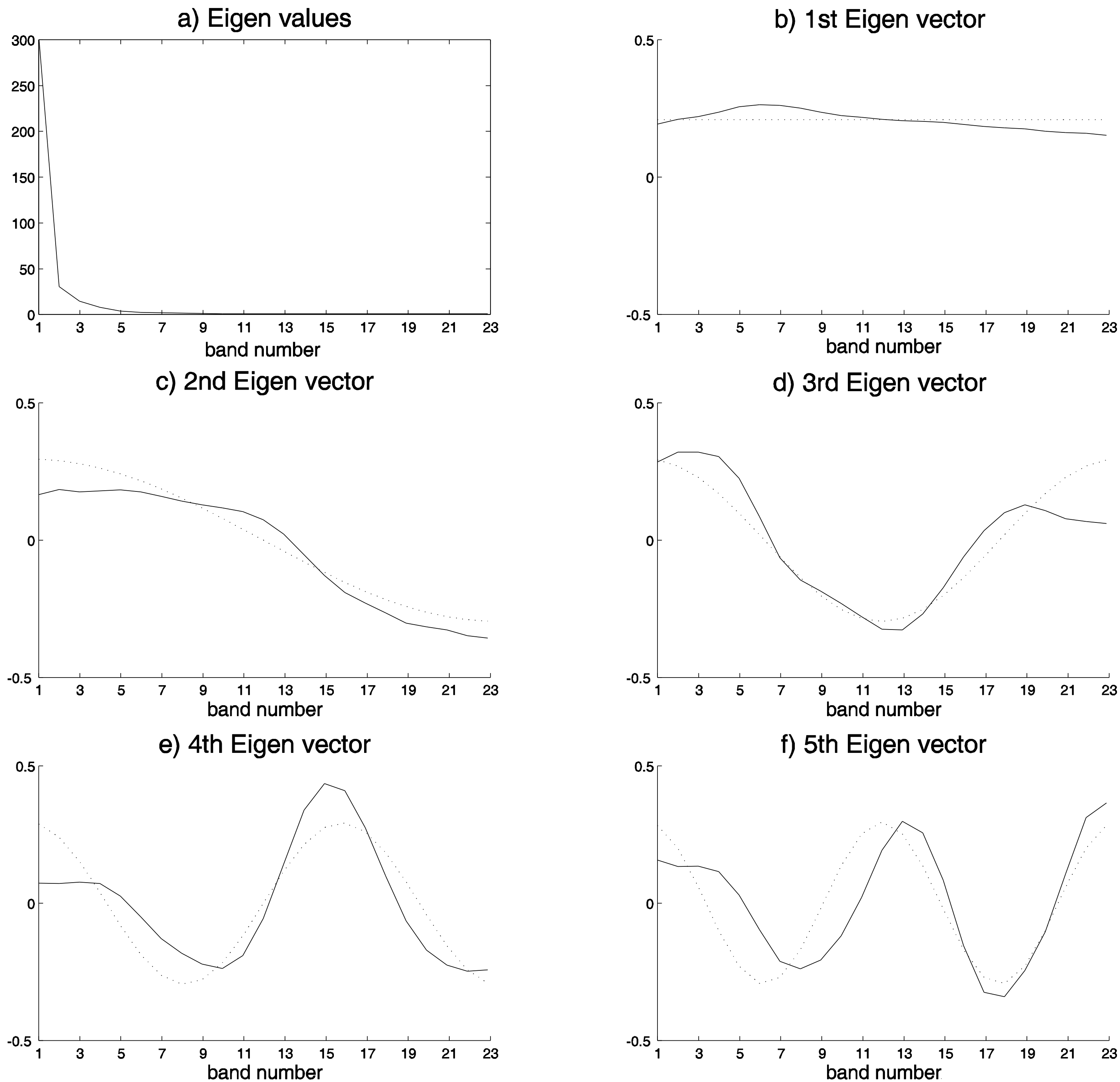

a) Segment of speech signal for vowel 'iy'
time [ms]
b) Speech segment after preemphasis and windowing
time [ms]
c) Fourier spectrum of speech segment
frequency [Hz]
d) Filter bank energies - smoothed spectrum
band number
e) Log of filter bank energies
band number
f) Mel frefuency cepstral coefficients
mel quefrency

Původní signál



Logaritmický vystup z banky filtru – je třeba již jen dekorelovat




a) Eigen values
300
250
200
150
100
50
0
1 3 5 7 9 11 13 15 17 19 21 23
band number
b) 1st Eigen vector
0.5
0
-0.5
1 3 5 7 9 11 13 15 17 19 21 23
band number
c) 2nd Eigen vector
0.5
0
-0.5
1 3 5 7 9 11 13 15 17 19 21 23
band number
d) 3rd Eigen vector
0.5
0
-0.5
1 3 5 7 9 11 13 15 17 19 21 23
band number
e) 4th Eigen vector
0.5
0
-0.5
1 3 5 7 9 11 13 15 17 19 21 23
band number
f) 5th Eigen vector
0.5
0
-0.5
1 3 5 7 9 11 13 15 17 19 21 23
band number

Singular Value Decomposition - SVD

$$A = UDV^T$$

- **A** je jakákoli mxn matice
- **U** je mxn matice kde sloupce jsou ortonormální báze
- **V** je nxn matice kde sloupce jsou ortonormální báze
- **D** je nxn je diagonální matice

- Předpokládejme, že matice A je matice s příznakovými vektory v řádcích s již odečtenou střední hodnotou ➔ **Σ** = $\mathbf{A}^T\mathbf{A}$
- Potom z následujících vztahů vyplývá, ze:
  - **V** jsou vlastní vektory **Σ**
  - Diagonála **D** obsahuje odmocniny z vlastních čísel **Σ** (variance ve směrech vlastních vektorů)

$$A^T A = VDU^T UDV^T = VD^2 V^T$$

$$AA^T = UDV^T VDU^T = UD^2 U^T$$