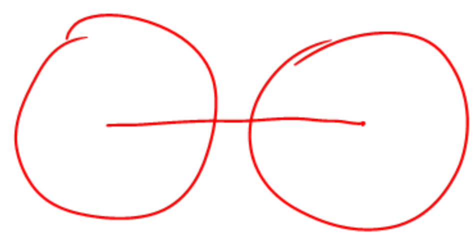
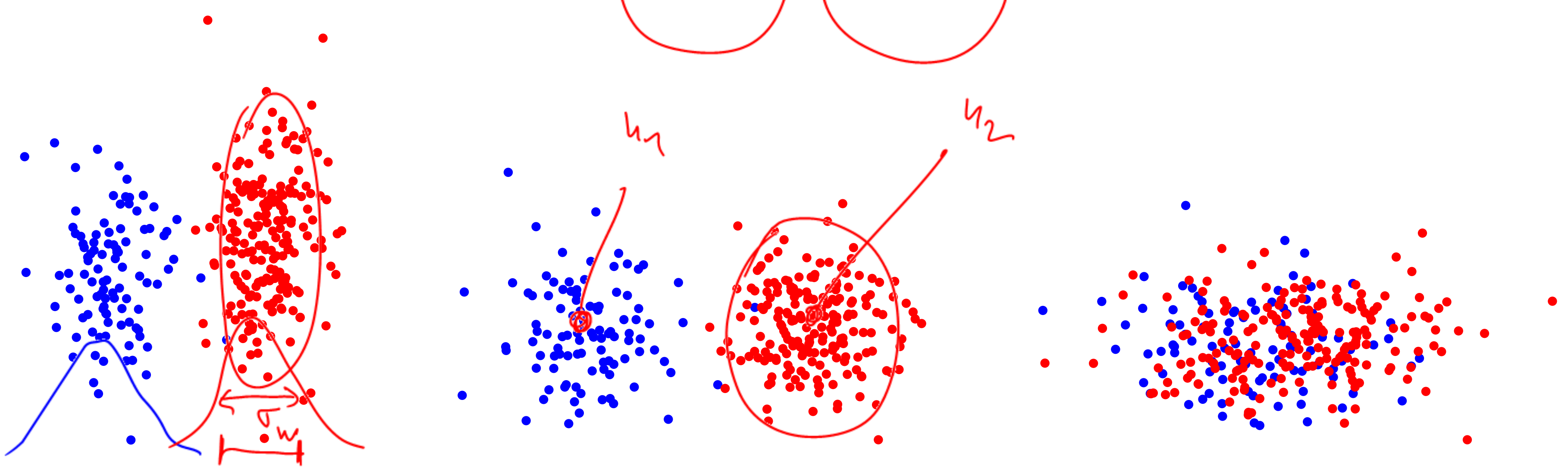
Fisherova lineární diskriminační analýza (FLDA)

Dána množina vektorů $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ v prostoru dimenze d, u každého identifiátor třídy ω_i, $i = 1, 2, \ldots, k$, $\omega_i \in \Omega$ (Ω je množina velikosti o), po sloupcích sestavíme $\mathbf{X}^\omega$ pro třídu $\omega \in \Omega$.

Nechť $\mathbf{W}^\top$ je matice $s \times d$, $s < d$, která redukuje dimenzi vektorů lineární transformací $\mathbf{Y} = \mathbf{W}^\top \mathbf{X}$. Budeme hledat takovou projekci, ve které jsou třídy co nejdále od sebe, měřeno vzhledem k rozptylu uvnitř tříd.



Př: body v dimenzi $d = 3$



projekce $x - y$ projekce $x - z$ projekce $y - z$

$\frac{\sigma_b}{\sigma_w} \to \max$

σ_b

FLDA: potlačení nepodstatných rysů, ať náhodných nebo ne

Úloha pro dvě třídy, $o = 2$

Promítáme $\mathbf{y} = \mathbf{w}^\top \mathbf{X}$, kde $\mathbf{y}$ je ‚matice' dimenze $1 \times k$.

Před projekcí jsou vektory středních hodnot

$$\mathbf{m}^\omega = \frac{1}{k^\omega} \sum_{i=1}^{k^\omega} \mathbf{x}_i^\omega,$$ kde k^ω je počet prvků ve třídě ω $m^\omega \in \mathbb{R}^d$

Vzdálenost mezi třídami po projekci měříme rozdílem středních hodnot

$\| w^T (m^1 - m^2) \|^2 = \| (m^1 - m^2)^T w \|^2 =$

$$|n^1 - n^2|^2 = |\mathbf{w}^\top \mathbf{m}^1 - \mathbf{w}^\top \mathbf{m}^2|^2 = \dots = \mathbf{w}^\top \underbrace{(\mathbf{m}^1 - \mathbf{m}^2)(\mathbf{m}^1 - \mathbf{m}^2)^\top}_{d \times d} \mathbf{w} \stackrel{\text{def}}{=} \mathbf{w}^\top \mathbf{S}_b \, \mathbf{w}$$

Rozptyl uvnitř tříd po projekci měříme jako

$$(s^1)^2 + (s^2)^2 = \sum_{i=1}^{k^1} \|\mathbf{w}^\top (\mathbf{x}_i^1 - \mathbf{m}^1)\|^2 + \sum_{j=1}^{k^2} \|\mathbf{w}^\top (\mathbf{x}_j^2 - \mathbf{m}^2)\|^2 = \dots = \mathbf{w}^\top (\mathbf{S}^1 + \mathbf{S}^2) \, \mathbf{w} \stackrel{\text{def}}{=} \mathbf{w}^\top \mathbf{S}_w \, \mathbf{w}$$

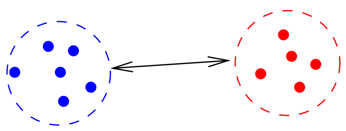
$\sum_i | \bar{x}_i^T w |^2 = \sum_i w^T \bar{x}_i \bar{x}_i^T w = w^T \Big(\underbrace{\sum \bar{x}_i \bar{x}_i^T}_{S} \Big) w$

pokračování

Fisherův diskriminant je

známo též jako Rayleigho poměr

$$F = \frac{|n^1 - n^2|^2}{(s^1)^2 + (s^2)^2} = \frac{\mathbf{w}^\top \mathbf{S}_b\, \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_w\, \mathbf{w}}.$$



Věta: Vektor $\mathbf{w}$ maximalizující F musí pro nějaké $\lambda \in \mathbb{R}$ vyhovovat podmínce

$$\frac{2 S_b w (w^T S_w w) - 2 S_w w (w^T S_b w)}{(w^T S_w w)^2} = 0 \rightarrow \quad S_b w = F S_w w \rightarrow (S_w^{-1} S_b) w = F \cdot w$$

$$\mathbf{S}_b\mathbf{w} = \lambda \mathbf{S}_w \mathbf{w}. \qquad (1)$$

- Nalezení takového $\mathbf{w}$ a λ se nazývá **zobecněný problém vlastních čísel**. Pokud je $\mathbf{S}_w$ regulární, lze ho vyřešit jako obyčejný problém $\mathbf{S}_w^{-1}\mathbf{S}_b\, \mathbf{w} = \lambda \mathbf{w}$.
- λ je hodnota F v optimu $\Rightarrow$ vybereme $\mathbf{w}$ odpovídající největšímu λ
- Pro $o = 2$ není třeba řešit (1), protože $\mathbf{S}_b\mathbf{w}$ je vždy paralelní s vektorem $\mathbf{m}^1 - \mathbf{m}^2$, takže

$$\mathbf{w} = \mathbf{S}_w^{-1}(\mathbf{m}^1 - \mathbf{m}^2)$$

  normalizovaný na jednotkovou délku

  důkaz: $\circledast$1
- Pokud je málo měření a dimenze $\mathbf{S}$ je velká, tak je $\mathbf{S}_w$ singulární a problém (1) musíme řešit obecně.

  V Matlabu pak použijeme `[W,D]=eig(Sb,Sw)`, `D` obsahuje λ_i na diagonále a `W` obsahuje zobecněné vlastní vektory po sloupcích

Úloha pro více tříd, $o > 2$

Matice vnitrotřídního rozptylu zůstává $\mathbf{S}_w = \sum_{i=1}^{o} \mathbf{S}^i$.

Zavedeme vážené střední hodnoty

vážený průměr pro $\mathbf{S}_b$ lze použít i pro $o = 2$

$$\mathbf{m} = \frac{1}{k}\sum_{i=1}^{o} k^i\,\mathbf{m}^i \qquad \mathbf{S}_b = \sum_{i=1}^{o} k^i(\mathbf{m}^i - \mathbf{m})(\mathbf{m}^i - \mathbf{m})^\top$$

Platí $\sum_{i=1}^{k}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^\top = S_w + S_b$

Všechny projekce sdružíme maticovým zápisem

$$\mathbf{Y} = \mathbf{W}^\top\mathbf{X}, \qquad \text{kde } \mathbf{Y} \text{ je } s \times k,\ \mathbf{W}^\top \text{ je } s \times d \text{ a } \mathbf{X} \text{ je } d \times k \text{ matice}$$

, $s \times s$ regulární

$\det(A) \neq 0$

Pak je Fisherův diskriminant

$\det$ – měří objem ‚elipsoidu kovariance' po projekci

$W \mapsto WA$

$$F = \frac{\det(\overbrace{\mathbf{W}^\top\mathbf{S}_b\mathbf{W}}^{s \times s})}{\det(\mathbf{W}^\top\mathbf{S}_w\mathbf{W})} = \frac{\det(A^\top W^\top S_b W A)}{\det(\quad S_w \quad)}$$

Věta: Sloupce $\mathbf{W}$ jsou tvořeny zobecněnými vlastními vektory odpovídajícími s největším zobecněným vlastním číslům $\mathbf{S}_b\mathbf{w}_i = \lambda_i\mathbf{S}_w\mathbf{w}_i,\ i = 1, 2, \ldots, s$ a v optimu platí $F = \prod_{i=1}^{s}\lambda_i$

Kanonické korelace

Pozorování: Pokud $\mathbf{W}$ je matice maximalizující F, potom $\mathbf{WA}$ také maximalizuje F, kde $\mathbf{A}$ je regulární $s \times s$ matice.

- Kanonická volba z této třídy je normalizovaná matice

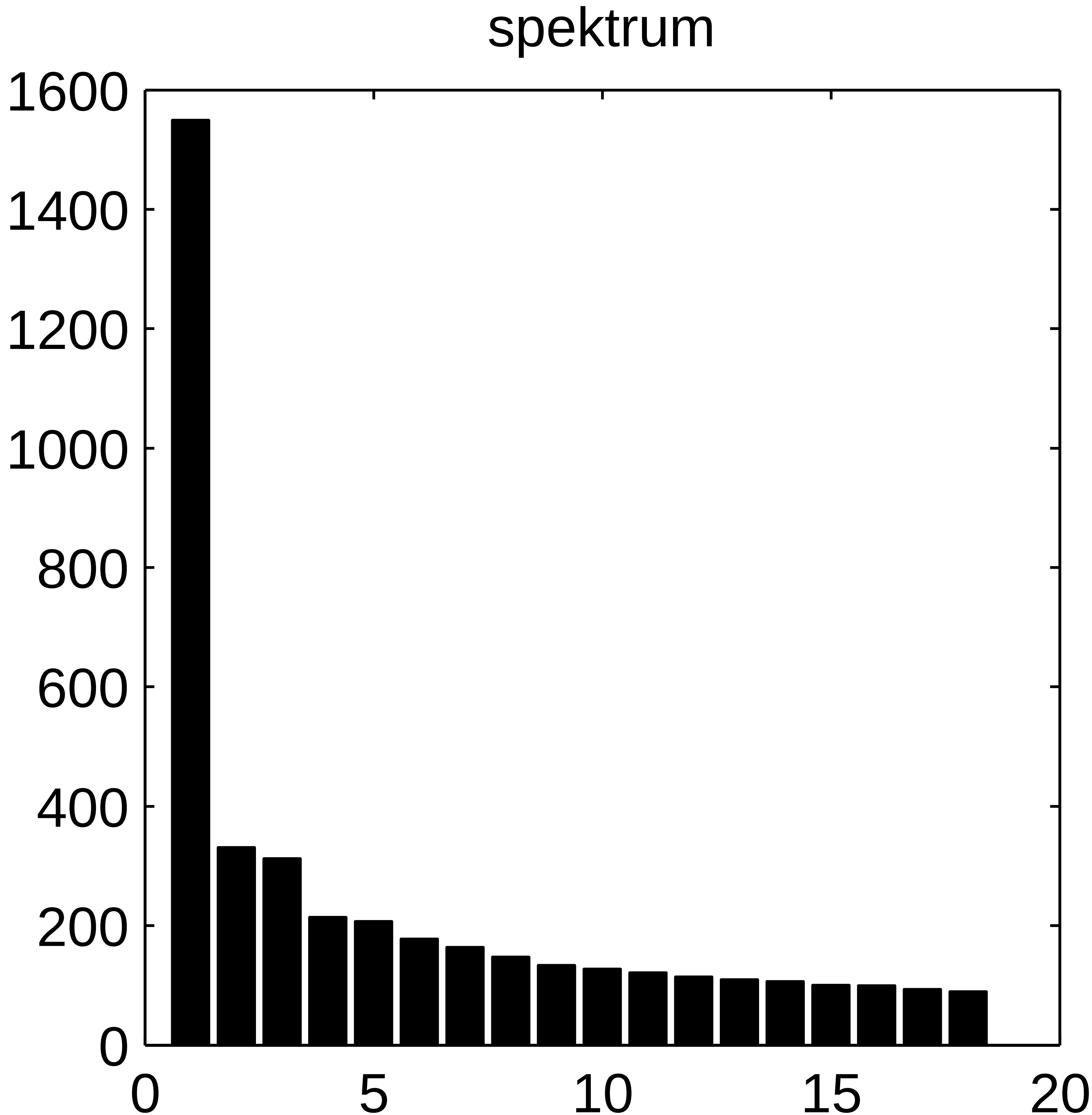
$$\mathbf{W}_0 = \mathbf{W} \left(\mathbf{W}^\top \mathbf{S}_w \mathbf{W}\right)^{-\frac{1}{2}}$$

- kovarianční matice mezi původním měřením x a jeho projekcí je $\mathbf{W}_0^\top \mathbf{S}_w$, kterou můžeme normalizovat na kanonické korelace

$$\mathbf{C} = \mathbf{W}_0^\top \mathbf{S}_w \operatorname{diag}(\mathbf{S}_w^{-\frac{1}{2}})$$

- notace: $\operatorname{diag} \mathbf{A}$ je diagonální matice získaná vynulováním nediagonálních prvků $\mathbf{A}$
- prvky $\mathbf{C}$ jsou v intervalu $\langle -1, 1 \rangle$
- kanonické korelace vypovídají o tom, kde je informace rozlišující třídy (které složky měření $\mathbf{x}$ jsou informativní)

Konec


spektrum
1600
1400
1200
1000
800
600
400
200
0
0
5
10
15
20