Pattern Recognition HW 3

171240510 Yuheng Ma

Math Major, Kuang Yaming Honors School

Spring 2020

1 06 FLD 3

1. $A^T\left(A^{-1}\right)^T=\left(A^{-1}A\right)^T=E$, thus $inv(X)'=inv(X')$, AA' and $A'A$ have same eigenvalues, thus $(inv(X)'inv(X))$ is similar to $((inv(X)inv(X)^T)=inv(X)inv(X')=(X'X)^{-1}$. Thus $\kappa_2(X)=\frac{\sigma_1}{\sigma_n}$.

2. By $Ax=b$, $A\Delta x=\Delta b$ and $\|b\|=\|Ax\|\leqslant\|A\|\|x\|$, $||\Delta x||=||A^{-1}\Delta b||\leqslant\|\mathrm{A}^{-1}\|\|\Delta\mathrm{b}\|$. $\frac{||\Delta x||}{||x||}\leqslant||A^{-1}||||A||\frac{||\Delta b||}{||b||}$. A large condition number will loss its control of error of x.

3. If A is orthogonal matrix, $||A||=||A^{-1}||=1$. Thus $\kappa(A)=1$.

2 06 FLD 6

1. .

2. .

3. PCA is unsupervised learning while FLD is supervised, which provides different type of discriminate method. This also results in difference in dims of eigenspace. Eigenface generally has a larger dim.

4. Around 300.

3 07 SVM 1

1. .

2. (a) 66.925% (b) Scale should be [0,1]. (c) 95.675% (d)87.7%.(e) C=1048.0,$\gamma=5.04$.

3. The first data set a1a is inbalenced with 395 positive class and 1210 negative class. It did not significantly improve something.

4 08 Probabilistic 2

1. $\int_{x_m}^{\infty}\frac{c}{x^{a+1}}dx=1$ thus $\frac{c}{\alpha}x_m^{-\alpha}=1\Rightarrow c=\alpha x_m^{\alpha}$.

2.

$$
\begin{aligned}
&l\left(x_1,\ldots x_n\right)\\
&=\prod_{i=1}^n\frac{\alpha x_m^{\alpha}}{x_i^{\alpha+1}}\\
&=(\alpha+1)\sum\log x_i+\alpha\cdot n\ln\alpha+n\alpha\log x_m
\end{aligned}
$$

By taking partial derivative, $\hat{x_m}=max(x_1,\cdots,x_n)$, $\hat{\alpha}=\frac{1}{n\log\prod\frac{x_i}{\hat{x_m}}}$

3. $$\begin{aligned} P(\theta|x) &\propto p(x|\theta)P(\theta|x_m,k) \\ &= \begin{cases} \left(\frac{1}{\theta}\right)^n & \theta \geqslant \max(x_i) \\ 0 & \theta < \max(x_i) \end{cases} \begin{cases} \frac{kx_m^k}{\theta^{k+1}} & \theta \geqslant x_m \\ 0 & \theta < x_m \end{cases} \\ &\propto \begin{cases} \frac{1}{\theta^{n+k+1}} & \theta \geqslant \min(\max(x_i), x_m) \\ 0 & \text{others} \end{cases} \end{aligned}$$

   Thus $\alpha' = n+k+1$ and $x_m = \min(\max(x_i), x_m)$.

5 09 Metric 6

1. .
2. Around 85%.
3. Around 65%.
4. Every feature is push to 1 thus harder to discriminate.

6 10 IT 2

1. d(x,y)=d(y,x), d(x,x)≥0 and $d(x,x) = 0 \Leftrightarrow x = \theta$, d(x+y)+d(y+z)≥d(x+z).
2. No.

$$KL(A||A) = 0 \tag{1}$$
$$KL(A||B) = \frac{1}{2}\log\frac{2}{3} \tag{2}$$
$$KL(B||A) = \frac{1}{4}\log\frac{31}{8} \tag{3}$$
$$KL(B||C) = \frac{1}{4}\log\frac{2274}{2058} \tag{4}$$
$$KL(A||C) = \frac{1}{2}\log\frac{32}{7} \tag{5}$$

   (1) rejects the second, (2)(3) reject the first, (2)(4)(5) reject the last.

3.
```
import numpy as np
def KL(x,y):
return x[1]*(np.log(x[1])-np.log(y[1]))+x[0]*(np.log(x[0])-np.log(y[0]))
A=np.array([0.5,0.5])
B=np.array([0.25,0.75])
C=np.array([0.125,0.875])
print(KL(A,A),KL(A,B)-KL(B,A),KL(A,B)+KL(B,C)-KL(A,C))

out:0.0 0.013029000284753484 -0.2118244650968008
```

7 10 IT 6

$$h(X) = -\int_0^\infty \lambda e^{-\lambda x}(\ln\lambda + -\lambda x)dx$$

And

$$\begin{aligned} 0 \leq& KL(p\|q) \\ &= \int p(x)\ln\frac{p(x)}{q(x)} \\ &= \int p(x)\ln p(x) - \int p(x)\ln q(x) \end{aligned}$$

and

$$
\begin{aligned}
&\int p(x) \log q(x) \\
=&\int p(x)(\log \lambda - \lambda x) dx \\
=&\log \lambda - 1
\end{aligned}
$$

Thus proved.