模式识别作业3

张祎扬 181840326 匡亚明学院本科生

1.讲义6习题3

(a)$\kappa_2(X)=\sqrt{\frac{\lambda_{max}(X)}{\lambda_{min}(X)}}=\frac{\sigma_{max}(X)}{\sigma_{min}(X)}=\frac{\sigma_1}{\sigma_n}$

(b)假设A精确，b有误差δb

$$A(\delta x+x)=b+\delta b$$
$$beacuse\ \ Ax=b\ \ ,so\ \ \delta x=A^{-1}\delta b$$
$$because\ \ ||\delta x||\leq||A^{-1}||||\delta b||\ and\ ||b||=||Ax||\leq||A||||x||$$
$$so\ \frac{||\delta x||}{||x||}\leq||A^{-1}||||A||\frac{||\delta b||}{||b||}=\kappa_2(A)\frac{||\delta b||}{||b||}$$

说明右端项的相对误差$\frac{||\delta b||}{||b||}$在解中可能放大了$\kappa_2(A)$倍.

假设b精确，A有误差δA

$$(A+\delta A)(x+\delta x)=b\quad\delta x=-A^{-1}\delta A(x+\delta x)$$
$$so\ \ \frac{||\delta x||}{||x+\delta x||}\leq||A^{-1}||||\delta A||=||A||||A^{-1}||\frac{||\delta A||}{||A||}=\kappa_2(A)\frac{||\delta A||}{||A||}$$

说明矩阵A的相对误差$\frac{||\delta A||}{||A||}$在解中可能放大了$\kappa_2(A)$倍.

所以如果$\kappa_2(A)$很大，说明该线性系统是病态的.

(c)正交矩阵有性质$A^TA=I$, 所以$\kappa_2(A)=||A||||A^{-1}||=\sqrt{\frac{\lambda_{max}(A^TA)}{\lambda_{min}(A^TA)}}=\sqrt{\frac{\lambda_{max}(I)}{\lambda_{min}(I)}}=1$. 条件数较小，所以正交矩阵是良态的。

2.讲义6习题6

(a)ORL人脸数据集共包含40个不同人的400张图像,此数据集下包含40个目录，每个目录下有10张图像，每个目录表示一个不同的人。所有的图像是以PGM格式存储，灰度图，图像大小宽度为92，高度为112。对每一个目录下的图像，这些图像是在不同的时间、不同的光照、不同的面部表情(睁眼/闭眼，微笑/不微笑)和面部细节(戴眼镜/不戴眼镜)环境下采集的。所有的图像是在较暗的均匀背景下拍摄的，拍摄的是正脸(有些带有略微的侧偏)。
参考原文链接：https://blog.csdn.net/fengbingchun/java/article/details/79008891

(b)已下载和学习

(c)PCA的识别效果比FLD强一些，因为PCA的目的是最大化方差，不考虑类别，所以PCA所求得的特征值也相对较大。FLD考虑类别，目标是尽可能分开，并不是简单地最大化方差，效果不如PCA。

(d)不断增加eigenfaces,发现当eigenfaces的数量大概为320左右，重构的人脸与原始输入图像难以区分。

3.讲义7习题1

(a)已了解

(b)

i.

```
 zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-train -s 0 -t 2 svmguide1.txt model
...*..*
optimization finished, #iter = 5371
nu = 0.606150
obj = -1061.528918, rho = -0.495266
nSV = 3053, nBSV = 722
Total nSV = 3053
 zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-predict svmguide1.t model output
Accuracy = 66.925% (2677/4000) (classification)
```

使用默认参数的准确率是66.925%

ii.

```
 ✘ zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-scale -l -1 -u 1 -s range1 svmguide1.txt > svmguide
1.scale
 zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-scale -r range1 svmguide1.t > svmguide1.t.scale
 zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-train svmguide1.scale
*
optimization finished, #iter = 496
nu = 0.202599
obj = -507.307046, rho = 2.627039
nSV = 630, nBSV = 621
Total nSV = 630
 zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-predict svmguide1.t.scale svmguide1.scale.model svmgui
de1.t.predict
Accuracy = 96.15% (3846/4000) (classification)
```

测试集的准确率是96.15%

iii.

```
optimization finished, #iter = 3509115
nu = 0.121917
obj = -376.234540, rho = 5.887607
nSV = 381, nBSV = 375
Total nSV = 381
 zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-predict svmguide1.t model output
Accuracy = 95.675% (3827/4000) (classification)
```

使用线性核的准确率为95.675%

iv.

```
 zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-train -s 0 -t 2 -c 1000 svmguide1.txt model
....*..*
optimization finished, #iter = 6383
nu = 0.000721
obj = -1114.038221, rho = -0.407723
nSV = 3001, nBSV = 0
Total nSV = 3001
 zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-predict svmguide1.t model output
Accuracy = 70.475% (2819/4000) (classification)
```

使用C=1000以及RBF核的准确率是70.475%

v.超参数C=2，$\gamma = 2$，准确率是96.875%

从这些实验中，我学到了超参数的选择，内核函数的选择，以及是否对特征进行规范化对模型的准确率有很大的影响。因此，如果想要尽可能地提高模型的预测准确率，那么选择合适的参数和方法就显得尤为重要。

(c)找到不平衡数据集svmguide3

```
zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-train -s 0 -t 2  svmguide3.txt model3
*
optimization finished, #iter = 535
nu = 0.452614
obj = -545.901031, rho = -0.985060
nSV = 570, nBSV = 552
Total nSV = 570
zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-predict svmguide3.t model3 output3
Accuracy = 2.43902% (1/41) (classification)
```

用默认参数训练准确率极低，只有2.43902%

```
✘ zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-train -s 0 -t 2 -w1 2 svmguide3.txt model3
*
optimization finished, #iter = 728
obj = -1014.019545, rho = -2.318946
nSV = 815, nBSV = 805
Total nSV = 815
zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-predict svmguide3.t model3 output3
Accuracy = 34.1463% (14/41) (classification)
zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-train -s 0 -t 2 -w1 4 svmguide3.txt model3
*
optimization finished, #iter = 903
obj = -1576.553413, rho = -3.694247
nSV = 1055, nBSV = 1046
Total nSV = 1055
zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-predict svmguide3.t model3 output3
Accuracy = 78.0488% (32/41) (classification)
zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-train -s 0 -t 2 -w1 8 svmguide3.txt model3
.
WARNING: using -h 0 may be faster
*
optimization finished, #iter = 1230
obj = -1871.652637, rho = -2.116360
nSV = 1074, nBSV = 1057
Total nSV = 1074
zhangyiyang@MacBook-Pro  ~/Downloads/libsvm-3.24  ./svm-predict svmguide3.t model3 output3
Accuracy = 100% (41/41) (classification)
```

当使用参数-wi时，赋予class 1不同的权重，可以发现随着权重的增加，准确率越来越高。2时准确率是34.1463%，4时是78.0488%，而当权重是8时准确率就已经达到了100%。由此可以看出，-wi参数在处理数据时非常有用，适当的参数值可以对准确率产生非常正面的影响。

4.讲义8习题2

(a)

$$\int_{-\infty}^{+\infty} p_1(x)dx = \int_{-\infty}^{+\infty} \frac{c_1}{x^{\alpha+1}}dx = c_1(\frac{1}{-\alpha})x^{\alpha}|_{x_m}^{+\infty} = \frac{c_1}{\alpha}x_m^{-\alpha} = 1$$

$$so \;\; c_1 = \alpha x_m^{\alpha}$$

$$so \;\; p_1(x) = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}[x \geq x_m]$$

所以X服从$Pareto(x_m, \alpha)$.

(b)

$$L(\alpha, x_m) = \prod_{i=1}^{n} p(x_i|\alpha, x_m)$$

$$1.when \ \ \exists x_k < x_m, L(\alpha, x_m) = 0$$

$$2.when \ \ x_k \geq x_m(\forall k), L(\alpha, x_m) \neq 0$$

$$lnL(\alpha, x_m) = \sum_{i=1}^{n} lnp(x_i|\alpha, x_m) = \sum_{i=1}^{n} ln\frac{\alpha x_m^{\alpha}}{x^{\alpha+1}} = nln\alpha + n\alpha lnx_m - (\alpha+1)\sum_{i=1}^{n} lnx_i$$

$$so \ \frac{\partial lnL(\alpha, x_m)}{\partial x_m} = \frac{n\alpha}{x_m} > 0$$

所以$L(\alpha, x_m)$随着x_m单调递增，当x_m取最大值时取最大值，又因为$x_m \leq x_k, \forall k$,所以$x_m = x_{min}$.

$$\frac{\partial lnL(\alpha, x_m)}{\partial \alpha} = \frac{n}{\alpha} + nlnx_m - \sum_{i=1}^{n} lnx_i = 0$$

$$so \ \alpha = \frac{n}{\sum_{i=1}^{n} lnx_i - nlnx_m} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} lnx_i - lnx_m}$$

所以当$x_m = x_{min}, \alpha = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} lnx_i - lnx_m}$时是最大似然估计。

(c)

$$p(\theta|x_m, k) = Pareto(x_m, k) = \frac{kx_m^k}{\theta^{k+1}}(\theta \geq x_m)$$

$$p(D|\theta) = \prod_{i=1}^{n}\frac{1}{\theta} = \frac{1}{\theta^n}$$

$$p(\theta|D) = \frac{p(D|\theta) * p(\theta|x_m, k)}{p(D)} = c\frac{1}{\theta^n}\frac{kx_m^k}{\theta^{k+1}}$$

$$because \ \int_{-\infty}^{+\infty} c\frac{kx_m^k}{\theta^{n+k+1}}d\theta = 1, so \ c = \frac{(n+k)x_m^n}{k}$$

$$so \ \ p(\theta|D) = \frac{(n+k)x_m^n}{k}\frac{1}{\theta^n}\frac{kx_m^k}{\theta^{k+1}} = \frac{(n+k)x_m^{n+k}}{\theta^{(n+k)+1}}$$

当$\theta < x_m$时，$p(\theta|x_m, k) = 0$,所以$p(\theta|D) = 0$。

综上，$p(\theta|D) = \frac{(n+k)x_m^{n+k}}{\theta^{(n+k)+1}}[\theta \geq x_m]$也是一个Pareto分布$Pareto(x_m, n+k)$.

5.讲义9习题6

(a)已学习

(b)

```
✘ zhangyiyang@MacBook-Pro  ~/Downloads/liblinear-2.30  ./predict mnist.t model output
Accuracy = 84.17% (8417/10000)
```

使用默认参数的准确率时84.17%.

(c)使用数据变换之后准确率变为91.7%，提高了。

(d)将优化表示为对偶形式的拉格朗日函数,

$$L(w,b,\alpha) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i^T x_j) + \sum_{i=1}^{N}\alpha_i$$

$$\min_{\alpha}\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i^T x_j) - \sum_{i=1}^{N}\alpha_i \quad s.t.\sum_{i=1}^{N}\alpha_i y_i = 0$$

当开根变换之后，优化目标的第一项变小，优化效果更好。

6.讲义10习题2

(a)d需要满足的4个条件是：

1. d(x,y) = d(y,x).　对称性
2. 0 $\leq$ d(x,y) < $+\infty$ 非负性
3. $d(x,y) = 0 \rightleftharpoons x = y$ 同一性
4. d(x,z) $\leq$d(x,y) + d(y,z) 三角不等式

(b)KL散度不是一个有效的距离度量。

$$KL(A||A) = \sum_x p(x)log_2 1 = 0$$
$$for\ the\ same\ reason\ , KL(B||B) = KL(C||C) = 0$$
$$KL(A||B) = \frac{1}{2}log_2 2 + \frac{1}{2}log_2\frac{2}{3} = \frac{1}{2}log_2\frac{4}{3} > 0$$
$$KL(A||C) = \frac{1}{2}log_2 4 + \frac{1}{2}log_2\frac{4}{7} = \frac{1}{2}log_2\frac{16}{7} > 0$$
$$KL(B||A) = \frac{1}{4}log_2\frac{1}{2} + \frac{3}{4}log_2\frac{3}{2} = \frac{1}{4}log_2\frac{27}{16} > 0$$
$$KL(B||C) = \frac{1}{4}log_2 2 + \frac{3}{4}log_2\frac{6}{7} = \frac{1}{4}log_2\frac{432}{343} > 0$$
$$KL(C||A) = \frac{1}{8}log_2\frac{1}{4} + \frac{7}{8}log_2\frac{7}{4} = \frac{1}{8}log_2\frac{7^7}{4^8} > 0$$
$$KL(C||B) = \frac{1}{8}log_2\frac{1}{2} + \frac{7}{8}log_2\frac{7}{6} = \frac{1}{8}log_2\frac{7^7}{2*6^7} > 0$$

1.不满足对称性。$KL(A||B) \neq KL(B||A)$.

2.满足非负性

3.满足$KL(A||A) = KL(B||B) = KL(C||C) = 0$.

4.不满足。$KL(A||B) + KL(B||C) < KL(A||C)$

(c)测试代码如下：

```
import math
a = [0.5,0.5]
b = [0.25,0.75]
c = [0.125,0.875]
def KL(x,y):
    sum=0
```

```
    for i in range(0,len(x)):
        sum += x[i]*math.log2(x[i]/y[i])
    return sum
#测试交换律
if KL(a,b) == KL(b,a) and KL(a,c) == KL(c,a) and KL(b,c) == KL(c,b):
    print("符合性质1对称性")
else:
    print("不符合对称性")
#测试非负性
list = [a,b,c]
flag = 1
for i in range(0,len(list)):
    for j in range(0,len(list)):
        if KL(list[i],list[j])<0:
            flag = 0;
if flag == 0:
    print("不符合非负性")
else:
    print("符合非负性")
#测试性质3
sign = 1
for i in range(len(list)):
    if KL(list[i],list[i]) != 0:
        sign = 0
if sign == 0:
    print("不符合同一性")
else:
    print("符合同一性")
#测试三角公式
ok = 1
for i in range(len(list)):
    if KL(list[(i-1)%3],list[i]) + KL(list[i],list[(i+1)%3]) < KL(list[(i-1)%3],list[(i+1)%3]):
        ok = 0
if ok == 0:
    print("不符合三角不等式")
else:
    print("符合三角不等式")
```

输出的结果如下:

```
/usr/local/bin/python3 /Users/zhangyiyang/workspace/python_practice/patternreg.py
不符合对称性
符合非负性
符合同一性
不符合三角不等式
```

7.讲义10习题6

设$p(x)$是满足条件的指数分布，则$p(x) = \lambda e^{-\lambda x}(x \geq 0)$

$$h_p(X)=-\int_0^{+\infty}p(x)lnp(x)dx=-\int_0^{+\infty}\lambda e^{-\lambda x}ln(\lambda e^{-\lambda x})dx=-ln\lambda+1$$

设$q(x)$是任意分布

$$-\int_0^{+\infty}q(x)ln(p(x))dx=-\int_0^{+\infty}q(x)ln(\lambda e^{-\lambda x})dx=-\int_0^{+\infty}q(x)(ln\lambda-\lambda x)dx$$
$$=-ln\lambda\int_0^{+\infty}q(x)dx+\int_0^{+\infty}q(x)\lambda xdx=-ln\lambda*1+\lambda\int_0^{+\infty}xq(x)dx=-ln\lambda+\lambda E(X)$$
$$=-ln\lambda+\lambda\mu=-ln\lambda+1$$

令$f(x)=h_q(x)-h_p(x)$

$$f(x)=-\int_0^{+\infty}q(x)ln(q(x))dx+\int_0^{+\infty}p(x)ln(p(x))dx$$
$$=-\int_0^{+\infty}q(x)ln(q(x))dx+\int_0^{+\infty}q(x)ln(p(x))dx$$
$$=\int_0^{+\infty}q(x)ln\frac{p(x)}{q(x)}dx\le\int_0^{+\infty}q(x)(\frac{p(x)}{q(x)}-1)dx=\int_0^{+\infty}p(x)dx-\int_0^{+\infty}q(x)dx=1-1=0$$

所以，对于任意分布$q(x)$, $h_q(x)\le h_p(x)$,也就是说，参数为$\lambda=\frac{1}{\mu}$的指数分布是在这样约束条件的最大熵分布。