Pattern Recognition Exercise 3

张素威，MF1733082

May 2018

1 讲义6习题3

1.a

σ_1/σ_n

1.b

对矩阵$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix}$，用octave计算出的条件数为$4.0002e + 04$，这个值很大，计算$Ax = [2, 2]^T$，解得$x = [2, 0]^T$，计算$Ax = [2, 2.0001]^T$，解得$x = [1, 1]^T$，可见对$Ax = b$来说，$b$的0.0001的变化对$x$的影响很大，这里的$A$就是ill-conditioned。

1.c

对正交矩阵A来说，$AA^T = A^T A = E$，其奇异值都为1，故条件数也为1。所以正交矩阵是well-conditioned。

2 讲义6习题6

2.c

```
140: ./opencv/facerec_eigenfaces facerec.csv opencv/eigenfaces/
Predicted class = 39 / Actual class = 39.
Eigenvalue #0 = 2821437.63319
Eigenvalue #1 = 2061768.68526
Eigenvalue #2 = 1097036.82426
Eigenvalue #3 = 892072.93331
Eigenvalue #4 = 819432.79582
Eigenvalue #5 = 539015.30950
Eigenvalue #6 = 390872.85487
Eigenvalue #7 = 373689.19757
Eigenvalue #8 = 313679.14058
Eigenvalue #9 = 289046.71201
```

Figure 1: PCA运行截图

```
141: ./opencv/facerec_fisherfaces facerec.csv opencv/fisherfaces/
Predicted class = 39 / Actual class = 39.
Eigenvalue #0 = 67239.08797
Eigenvalue #1 = 4713.30259
Eigenvalue #2 = 1832.97376
Eigenvalue #3 = 1497.24676
Eigenvalue #4 = 631.96906
Eigenvalue #5 = 377.93553
Eigenvalue #6 = 256.68486
Eigenvalue #7 = 197.19217
Eigenvalue #8 = 148.88811
Eigenvalue #9 = 115.20196
Eigenvalue #10 = 85.59302
Eigenvalue #11 = 79.08640
Eigenvalue #12 = 56.77618
Eigenvalue #13 = 46.21607
Eigenvalue #14 = 40.12289
Eigenvalue #15 = 33.86131
```

Figure 2: FLD运行截图

从图中结果可以看出，两者都将最后一张图的类别分类正确，但是PCA得到的eigenvalue远大于FLD得到的特征值，这反应出了PCA和FLD的主要区别，PCA是无监督学习，只会选择大的特征值，FLD则是同时考虑了类间距离和类内距离，在最大化类间距离的同时最小化类内距离，这就使得特征值大小上有所妥协。

2.d

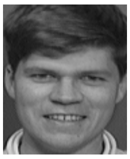

Figure 3: Original

Figure 4: 10

Figure 5: 40

Figure 6: 70

Figure 7: 100

Figure 8: 130

Figure 9: 160

Figure 10: 190

Figure 11: 220

Figure 12: 250

Figure 13: 280

从图中可以看出，使用220个特征值时，人脸已经很清晰了，使用250个特征值时很难看出和原图的差别。

3 讲义7习题1

3.b

(i) 66.925%

(ii) 96.15%(使用[-1,1]scale)

(iii) 95.675%(scale后的数据)

(iv) 96.725%(scale后的数据)

(v) 96.95%

对数据集进行合理的scale在svm中起着重要作用，往往能使得结果有飞跃性的进展，数据缩放后的调参只能在局部寻找一个最优的解。

3.c

数据集是svmguide3，未使用-wi参数时结果为85.3659%，使用参数-w-1 1 -w1 3时结果为100%，所以说设置-wi是有用的。

4 讲义8习题2

4.a

根据概率密度函数积分为1的公式得:

$$\int_{x_m}^{+\infty} \frac{c_1}{x^{\alpha+1}} dx = 1$$

解得$c_1 = \alpha x_m^{\alpha}$。

所以说X服从Pareto(x_m, α)分布。

4.b

将x_i代入，似然函数为

$$\frac{\alpha^n x_m^{n\alpha}}{(\prod_{i=1}^{n} x_i)^{\alpha+1}} [\![min\{x_i\} \geq x_m]\!]$$

指示函数左边的式子关于x_m单调增。所以x_m的极大似然估计就是

$$min\{x_i, 1 \leq i \leq n\}$$

对似然函数左半部取对数求导使为0:

$$\frac{n}{\alpha} + n \ln x_m - \sum_{i=1}^{n} \ln x_i = 0$$

则α的极大似然估计为$\frac{1}{\frac{\sum_{i=1}^{n} \ln x_i}{n} - \ln x_m}$。

4.c

由贝叶斯估计公式

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

其中$p(\theta) = \text{Pareto}(x_m, k)$，代入求得:

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

证毕！后验分布服从Pareto$(x_m, k+n)$。

5 讲义9习题6

5.b

84.17%

5.c

87.96%

5.d

简单猜想是因为开方运算是非线性运算，对特征开方相当于用了一个简单的核函数进行空间变换，使得结果精读更高。

6 讲义10习题2

6.a

对于任意x, y, z, d需要满足：
非负性：$d(x,y) \geq 0$
自反性: $d(x,y) = 0$ 当且仅当$x = y$
对称性：$d(x,y) = d(y,x)$
三角不等式：$d(x,z) \leq d(x,y) + d(y,z)$

6.b

KL散度不满足距离指标的定义要求。KL散度的定义公式为：

$$D(P||Q) = \sum_{x \in X} P(x) log \frac{P(x)}{Q(x)}$$

计算A, B, C中任意两种分布之间的KL散度：

$$D(A||B) = \frac{1}{2}log6 > 0, \quad D(B||A) = \frac{3}{4}log3 + log2 > 0$$

$$D(A||C) = 2log2 - \frac{1}{2}log7 > 0, \quad D(C||A) = \frac{7}{8}log7 - 4log2 > 0$$

$$D(B||C) = \frac{1}{4}log2 + \frac{3}{4}log\frac{6}{7} < 0, \quad D(C||B) = -\frac{1}{8}log2 + \frac{7}{8}log\frac{7}{6} > 0$$

显然，KL散度不满足非负性和对称性。
又因为$D(B||A) > D(B||C) + D(C||A)$,所以KL散度不满足三角不等式.
不过KL散度满足自反性.

6.c

```
import numpy as np
A = np.array([])
B = np.array([])
C = np.array([])
total = [A, B, C]
M = np.zeros(shape = (len(total), len(total)))
for i in range(len(total)):
  for j in range(len(total)):
    P = total[i]
    Q = total[j]
    M[i, j] = np.sum(P*np.log(P/Q))
print(M)
```

7 讲义10习题6

根据题目描述，可以建模出一个求最大熵的优化问题：

$$\max_q \int_{-\infty}^{+\infty} -q(x)\ln q(x)dx$$

$$s.t. \int_{-\infty}^{+\infty} xq(x)dx = \mu > 0$$

$$\int_{-\infty}^{+\infty} q(x)dx = 1$$

采用拉格朗日乘子法求解，该优化问题的对偶问题写作：

$$J(q) = \int_{-\infty}^{+\infty} q(x)\ln q(x)dx - \alpha_0(\int_{-\infty}^{+\infty} q(x)dx - 1) - \alpha_1 \int_{-\infty}^{+\infty} xq(x)dx$$

求偏导数：

$$\frac{\delta J}{\delta q}(q) = \ln(q) + 1 - \alpha_0 - \alpha_1 x = 0$$

求得：

$$q(x) = e^{\alpha_0 - 1} e^{\alpha_1 x}$$

将$q(x)$代入第二个约束方程可得：

$$e^{\alpha_0 + 1} = -\alpha_1$$

$q(x)$可以写作：

$$q(x) = \lambda e^{-\lambda x}$$

根据随机变量的期望约束可以求得：

$$\lambda = \frac{1}{\mu}$$

即当连续型随机变量服从参数为$\frac{1}{\mu}$的指数分布时，其熵才能达到最大值。