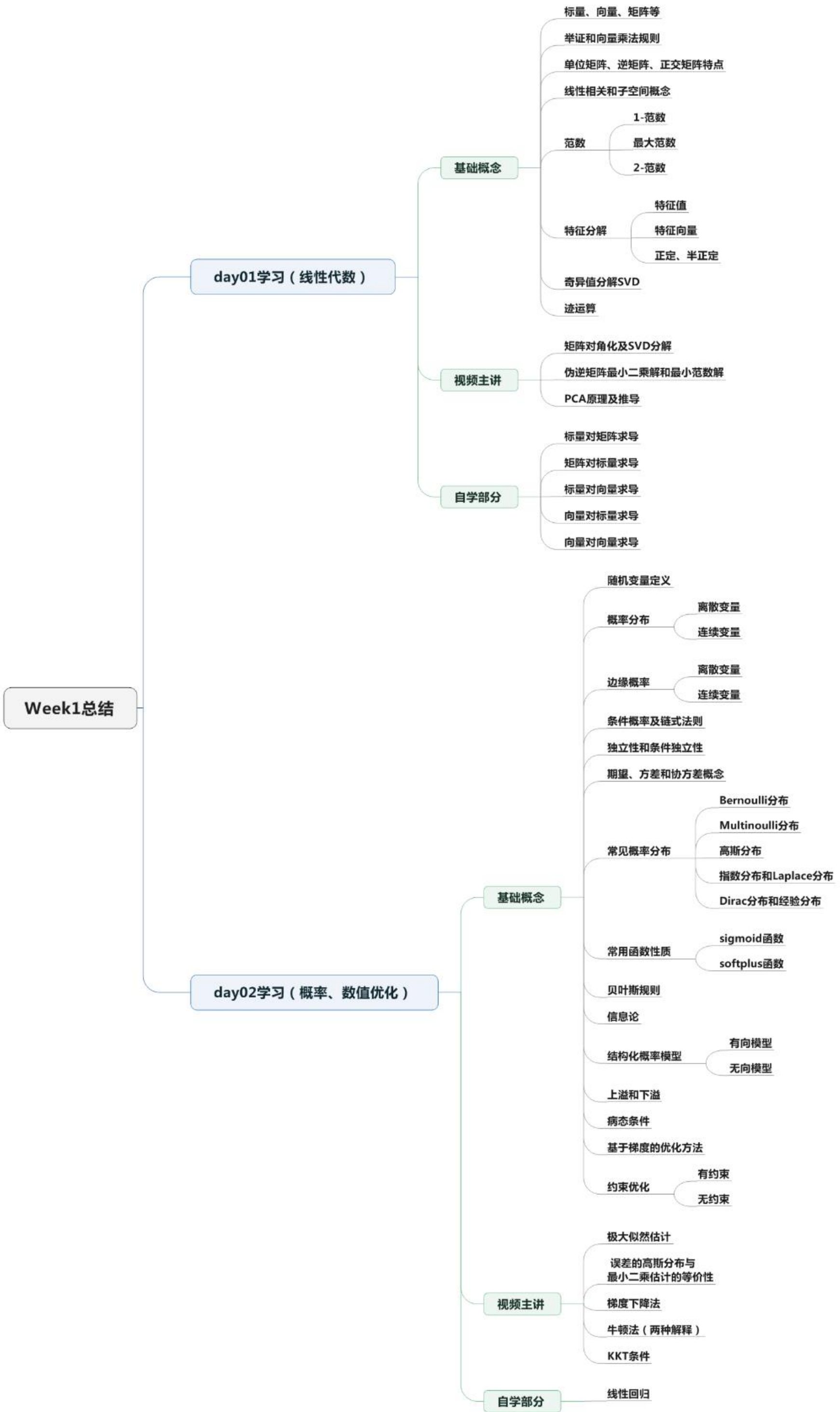
《深度学习》花书学习

第一章 Week01 总结

1.1 重点知识



1.2 深度学习最新发展

1. 计算机视觉

   生物特征识别：人脸识别、步态识别、行人 ReID、瞳孔识别；

   图像处理：分类标注、以图搜图、场景分割、车辆车牌、OCR、AR；

   视频分析：安防监控、智慧城市；

2. 自然语言处理

   语音识别、文本数据挖掘、文本翻译；

3. 数据挖掘

   消费习惯、天气数据、推荐系统、知识库（专家系统）；

4. 游戏

   角色仿真、AlphaGo（强化学习）、腾讯“觉悟”；

5. 复合应用

   无人驾驶、无人机、机器人；

6. 工业界

   机器故障预测、机械手控制；

1.3 主成分分析推导

方式 1：最大化方差

$$\begin{aligned} E(x\cdot u - E(x\cdot u))^2 &= E[(u\cdot(x-Ex))^2] \\ &= uE[(x-Ex)\cdot(x-Ex)^T]u \\ &= u^T C u \end{aligned}$$

方式 2：最小化误差

Find PCA basis vectors u that minimize $E||x - \hat{x}||^2$ for a partial expansion out to P components:

$$\hat{x} = \sum_{k=1}^{P} (x \cdot u_k) u_k$$

$$x - \hat{x} = \sum_{k=P+1}^{N} (x \cdot u_k) u_k$$

where N is the full set of vectors necessary to represent the data.

So, minimize the square of the last sum. The cross terms disappear because of the orthogonality of u_k. For each term:

$$E((x \cdot u)u)^2 = Eu(x \cdot u)(x \cdot u)u$$

the outer u's disappear because $u \cdot u = 1$.

$$= E(x \cdot u)(x \cdot u) = uCu$$

But $uCu = \lambda$, so the truncation error is the sum of the lower eigenvalues! Why: we know that u are eigenvectors, so they satisfy $Cu = \lambda u$, also $u \cdot u = 1$, so....

方式 3：对角化 相关矩阵

For a vector x, Exx^T is a correlation matrix.

Say M is a matrix whose columns contain data vectors. I think both MM^T and MM^T can be interpreted as correlation matrices.

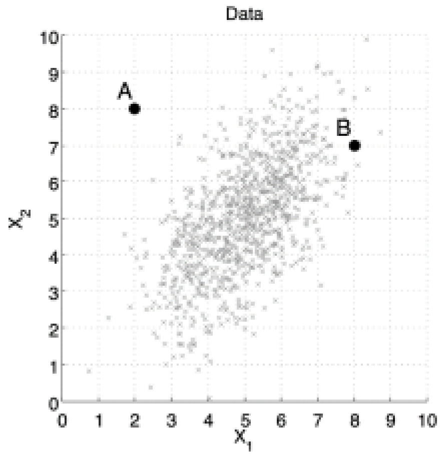
MM^T is the usual correlation matrix, a sum of outer products:

$$\begin{aligned} (MM^T)_{i,j} &= \sum_k x_k[i] x_k[j] \\ (MM^T) &= \sum_k x_k x_k^T \approx Exx^T = C \end{aligned}$$

If x_k are a sliding window through a signal, i.e. x_0 contains samples 0..10, x_1 samples 1..11, etc., then this corresponds to estimating the autocovariance of the signal. If x_k are images scanned into a vector, this gives the average (after dividing by N) correlation of pixel i with pixel j.

The i, j entry of $M^T M$ is the dot of data vector i with data vector j. If a column of M contains various measurements for a particular person then $(M^T M)_{i,j}$ gives the correlation, averaged across tests, of person i with person j, while $(MM^T)_{i,j}$ gives the correlation, averaged across people, of test i versus test j.

1.4 练习



1. 观察图中数据，估计该二元正态分布的均值μ（近似至最近的整数）。
2. 协方差矩阵Σ中，$\Sigma_{1,2} = Cov(X_1,X_2)$是正数？负数？零？
3. 图中数据的第一、二主成分的方向分别为协方差矩阵Σ的特征向量v_1和v_2，且$\|v_1\| = \|v_2\| = 1$。对于$X = [X_1\ X_2]^T$，这些方向定义了一组变换的基：

$$Z_1 = (X-\mu)^T v_1$$
$$Z_2 = (X-\mu)^T v_2$$

判断$\Sigma_{1,2} = Cov(Z_1,Z_2)$是正数？负数？零？

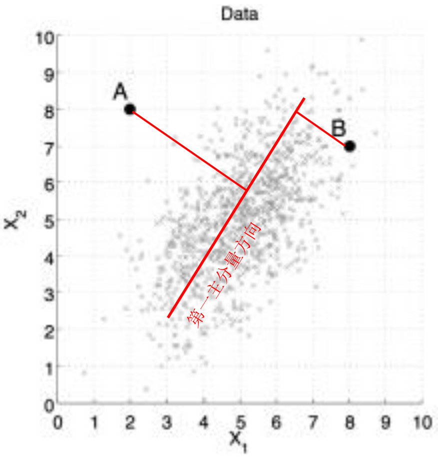
4. 在把数据投影到第一主成分方向v_1上后，A 和 B 中哪一点的 reconstruction error（原始数据与其在v_1上的投影间距离的平方）更高？

1、样本均值是总体均值的无偏估计，所以可以使用样本均值近似总体均值 μ，通过观察数据，可以发现 x1 和 x2 的均值大概均为 5，所以$\mu^T = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$。

2、$\Sigma_{1,2}$表示 x1 与 x2 的相关关系，通过数据可以发现，x1 与 x2 具有正相关关系，所以$\Sigma_{1,2} > 0$，所以为正数。

3、经过变换后，Z1 和 Z2 为 X 分别在 v1 和 v2 方向上的投影，而 v1 和 v2 是正交的，因而 Z1 和 Z2 是相互独立的，即$Cov(Z_1,Z_2)=0$。

4、第一主分量具有最大的特征值，而特征值代表在主分量上数据的方差大小，因而通过数据可以确定第一主分量的方向如下图所示，所以可以很直观地看出，A 点的重构误差更大。


Data
A
B
第一主分量方向
X_1
X_2
0
1
2
3
4
5
6
7
8
9
10