1. 部分预备知识

1.1. 一阶统计量

1. 均值：$E\{\vec{x}\} = \int_{-\infty}^{+\infty} \vec{x} p_x(\vec{x}) \, \mathrm{d}\vec{x}$ $p(\vec{x})$ 为 $\vec{x}$ 的概率密度函数

$$E\{\vec{g}(\vec{x})\} = \int_{-\infty}^{+\infty} \vec{g}(\vec{x}) p_x(\vec{x}) \, \mathrm{d}\vec{x}$$

性质：（1）$E\left\{\sum_{i=1}^{n} a_i \vec{x}^i\right\} = \sum_{i=1}^{n} a_i E\left\{\vec{x}^i\right\}$

（2）$E\{[A]\vec{x}\} = [A]E\{\vec{x}\}$

而在实际计算中，$E\{\vec{x}\} \approx \dfrac{1}{N}\sum_{p=1}^{N} \vec{x}^p$ $\quad E\{\vec{g}(\vec{x})\} \approx \dfrac{1}{N}\sum_{p=1}^{N} \vec{g}\left(\vec{x}^p\right)$

1.2. 二阶统计量

1. 自相关矩阵 $\vec{x} = [x_1, x_2, \cdots, x_n]^\top$

$$[R_x] = E\left\{\vec{x}\vec{x}^\top\right\} = E\left\{\begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2^2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_n^2 \end{bmatrix}\right\} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nn} \end{bmatrix} Y_{ij} = E\left\{x_i x_j^\top\right\}$$

性质：（1）$[R_x] = [R_x]^\top$

（2）$\vec{\alpha}^\top [R_x] \vec{\alpha} \geqslant 0 \quad \vec{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_n]^\top \neq 0$

（3）$[R_x]$ 的特征值 $\geqslant 0$

2. 自协方差矩阵

$$[C_x] = E\left\{(\vec{x} - E\{\vec{x}\})(\vec{x} - E\{\vec{x}\})^\top\right\} = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}$$

实际计算中 $[C_x] = \dfrac{1}{N}\sum_{i=1}^{N} \left(\vec{x}^p - \bar{\vec{x}}\right)\left(\vec{x}^p - \bar{\vec{x}}\right)^\top$ $\quad \bar{\vec{x}}$：均值

3. 互相关矩阵

$\vec{x} = [x_1, x_2, \cdots, x_n]^\top$ $\quad \vec{y} = [y_1, y_2, \cdots, y_n]^\top$ $\quad [R_{xy}] = E\left\{\vec{x}\vec{y}^\top\right\}$

4. 互协方差矩阵

$[C_{xy}] = E\left\{(\vec{x} - E\{\vec{x}\})(\vec{y} - E\{\vec{y}\})^\top\right\}$

1.3. 不相关、独立

1. 不相关

- 若 $[C_{xy}] = [0]$，则 $\vec{x},\vec{y}$ 不相关
- 若 $\vec{x},\vec{y}$ 不相关，则 $[R_{xy}] = E\left\{\vec{x}\vec{y}^\top\right\} = E\{\vec{x}\}E\left\{\vec{y}^\top\right\}$
- 若 $\vec{x}$ 各分量之间不相关

$$[C_x] = \begin{bmatrix} C_{11} & 0 & \cdots & 0 \\ 0 & C_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_{nn} \end{bmatrix} = \begin{bmatrix} \delta_{11}^2 & 0 & \cdots & 0 \\ 0 & \delta_{22}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_{nn}^2 \end{bmatrix} = diag\left\{\delta_{11}^2, \delta_{22}^2, \cdots, \delta_{nn}^2\right\}$$

- 若 $E\{\vec{x}\} = \vec{0}$，则 $[R_x] = [C_x] = [\mathrm{I}]$

2. 独立

- 若 $\rho_{xy}(\vec{x},\vec{y}) = \rho_x(\vec{x})\rho_y(\vec{y})$，则 $\vec{x}$ 与 $\vec{y}$ 相互独立
- $\vec{x}$ 和 $\vec{y}$ 相互独立时，则 $\vec{x}$ 和 $\vec{y}$ 一定不相关；反之不成立
- $\vec{x}$ 和 $\vec{y}$ 独立时，则 $E\{g(\vec{x})h(\vec{y})\} = E\{g(\vec{x})\}E\{h(\vec{y})\}$

2. 主元分析（PCA）

2.1. 原理

记 $\vec{x}^p \to \vec{y}^p \quad \vec{x}^p \in R^n \quad \vec{y}^p \in R^m\left(\text{一般 } m < n\right)$

对于 $\vec{x}^p \in R^n$，若找到一个新空间 U，其正交基为 $\vec{u}_1, \cdots, \vec{u}_n$，$\vec{u}_i^\top \vec{u}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

则在空间 U 中，$\vec{x}^p = a_1^p \vec{u}_1 + \cdots + a_i^p \vec{u}_i + \cdots + a_n^p \vec{u}_n = \sum_{k=1}^{n} a_k^p \vec{u}_k \quad ①$

其中 a_i^p 是 $\vec{x}^p$ 在 $\vec{u}_i$ 上的投影系数 $a_i^p = \vec{u}_i^\top \vec{x}^p = \vec{x}^{p\top} \vec{u}_i$

当 $m = n$ 时，$\vec{x}^p$ 可用 $\vec{u}_1, \cdots, \vec{u}_n$ 线性表示

当 $m < n$ 时，① 式中最大的 m 个 a_i^p 对应的 m 个主元作为对 $\vec{x}^p$ 的估计，估计值记作 $\vec{\tilde{x}}$

则可知 $\tilde{\vec{x}}^p = \sum_{i=1}^{m} a_i^p \vec{u}_i + \sum_{i=m+1}^{n} b_i \vec{u}_i$，其中 b_i 对所有样本采取同样的值

故可知实际值和估计值之间的误差为

$$E = \frac{1}{2}\sum_{p=1}^{N}\left(\vec{x}^p - \tilde{\vec{x}}^p\right)^\top\left(\vec{x}^p - \tilde{\vec{x}}^p\right) = \frac{1}{2}\sum_{p=1}^{N}\left(\sum_{i=m+1}^{n}\left(a_i^p - b_i\right)\vec{u}_i\right)^\top\left(\sum_{i=m+1}^{n}\left(a_i^p - b_i\right)\vec{u}_i\right)$$

化简可知 $E = \frac{1}{2}\sum_{p=1}^{N}\sum_{i=m+1}^{n}\left(a_i^p - b_i\right)^2$

若要使得误差 E 最小，则 $\frac{\partial E}{\partial b_i} = 0 \Rightarrow \sum_{p=1}^{N} b_i = \sum_{p=1}^{N} a_i^p$

假设 b_i 值相同，则 $N \cdot b_i = \sum_{p=1}^{N} a_i^p \Rightarrow b_i = \frac{1}{N}\sum_{p=1}^{N} a_i^p$

又因为 $a_i^p = \vec{u}^\top \vec{x}^p \Rightarrow b_i = \frac{1}{N}\sum_{p=1}^{N}\vec{u}_i^\top \vec{x}^p = \vec{u}_i^\top\left(\frac{1}{N}\sum_{p=1}^{N}\vec{x}^p\right) = \vec{u}_i^\top \overline{\vec{x}^p}$

代入 E 的表达式中得

$$E = \frac{1}{2}\sum_{p=1}^{N}\sum_{i=m+1}^{n}\left(\vec{u}_i^\top \vec{x}^p - \vec{u}_i^\top \overline{\vec{x}^p}\right) = \frac{1}{2}\sum_{i=m+1}^{n}\vec{u}_i^\top\left(\sum_{p=1}^{N}\left(\vec{x}^p - \overline{\vec{x}}\right)\left(\vec{x}^p - \overline{\vec{x}}\right)^\top\right)\vec{u}_i$$

又 $\because$ $[C_x] = N\sum_{p=1}^{N}\left(\vec{x}^p - \overline{\vec{x}}\right)\left(\vec{x}^p - \overline{\vec{x}}\right)^\top$ $\therefore$ $E = \frac{N}{2}\sum_{i=m+1}^{n}\vec{u}_i^\top[C_x]\vec{u}_i$

当 $\vec{u}_i$ 是 $[C_x]$ 的特征向量时，$[C_x]\vec{u}_i = \lambda_i \vec{u}_i, \lambda_i$ 为 $[C_x]$ 的特征根

则可知 $E = \frac{N}{2}\sum_{i=m+1}^{n}\lambda_i$

当对应于最大的 m 个特征根的 m 个特征向量，构成主分量，可使得误差 E 最小

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m \geqslant \cdots \geqslant \lambda_n$$

2.2. PCA 算法

已知样本 $\left\{\vec{x}^p\right\}_{p=1}^{N}, \vec{x}^p \in R^n$，降维至 $R^m, m \leqslant n$

- 求均值 $\overline{\vec{x}} = \frac{1}{N}\sum_{p=1}^{N}\vec{x}^p$

- 求协方差矩阵 $[C_x] = \frac{1}{N}\sum_{p=1}^{N}\left(\vec{x}^p - \vec{\bar{x}}\right)\left(\vec{x}^p - \vec{\bar{x}}\right)^\top$
- 求 $[C_x]$ 的特征值及特征向量
  $[C_x]$ 的 m 个最大的特征值对应的特征向量，记为 $\vec{u}_{maxi}(i = 1, 2, \cdots, m)$
- $y_i^p = a_i^p = \left(\vec{x}^p - \vec{\bar{x}}\right) \cdot \vec{u}_{maxi}$

$$\text{性质：}\begin{cases}\text{均值：}0\\ \text{方差：}\lambda_i\\ \vec{y}^p\ \text{各分量之间不相关}\end{cases}$$

例：图像压缩

1000 幅图像，每幅均是 512×512，原始数据量 $512 \times 512 \times 1000$

使用 PCA 进行压缩（取 20 个主元）后的数据量：$20 \times 512 \times 512 + 20 \times 1000 + 512 \times 512$

20×1000：1000 幅图像中每一副均有 20 个主元

$20 \times 512 \times 512$：20 个主元，对应的特征向量均为 512×512

512×512：均值

若以像素作为样本，数据量为 $20 \times 512 \times 512 + 20 \times 1000 + 1000$

2.3. 采用神经网络求取最大的特征向量

设 $\vec{u}_1$ 是最大特征值对应的特征向量，样本 $\vec{x}^p$（n 维，移除均值）

则可知 $y_1^p = a_1^p = \vec{x}^{p\top}\vec{u}_1 = \vec{u}_1^\top\vec{x}^p \quad \tilde{\vec{x}}^p = y_1^p\vec{u}_1 = \vec{u}_1 y_1^p$

若记 $\vec{u}_1 = \vec{\omega}_1$，则可知 $y_1 = \vec{\omega}_1^\top\vec{x} \quad \tilde{\vec{x}} = \vec{\omega}_1 y_1 = y_1\vec{\omega}_1$

目标：通过学习使权值 $\vec{\omega}_1$ 尽可能等于最大特征向量

2.3.1. 目标函数一

$$E(\vec{\omega}_1) = \frac{1}{2}\left|\left|\vec{x} - \tilde{\vec{x}}\right|\right|^2 = \frac{1}{2}||\vec{e}_1||^2 = \frac{1}{2}\vec{e}_1^\top\vec{e}_1$$

由梯度法得

$$\Delta\vec{\omega}_1 = -\eta\frac{\partial E(\vec{\omega}_1)}{\partial\vec{\omega}_1} = -\eta\frac{\partial\left(\frac{1}{2}\vec{e}_1^\top\vec{e}_1\right)}{\partial\vec{\omega}_1} = -\eta\frac{\partial\vec{e}_1}{\partial\vec{\omega}_1}\vec{e}_1 = -\eta\frac{\partial(\vec{x} - y_1\vec{\omega}_1)}{\partial\vec{\omega}_1}(\vec{x} - y_1\vec{\omega}_1)$$

$$
\begin{aligned}
\Delta\vec{\omega}_1 &= -\eta\left(-y_1[\mathrm{I}] - \vec{x}\vec{\omega}_1^\top\right)(\vec{x} - y_1\vec{\omega}_1) \\
&= \eta y_1(\vec{x} - y_1\vec{\omega}_1) + \eta\vec{x}\vec{\omega}_1^\top(\vec{x} - y_1\vec{\omega}_1) \\
&= \eta y_1(\vec{x} - y_1\vec{\omega}_1) + \eta\vec{x}y_1\left(1 - \vec{\omega}_1^\top\vec{\omega}_1\right)
\end{aligned}
$$

每次进行归一化处理 $\vec{\omega}_1 = \dfrac{\vec{\omega}_1}{||\vec{\omega}_1||}$，则可知 $\Delta\vec{\omega}_1 = \eta y_1(\vec{x} - y_1\vec{\omega}_1)$

则可知 $\vec{\omega}_1(t) = \vec{\omega}_1(t-1) + \eta y_1(t-1)(\vec{x}(t-1) - y_1(t-1)\vec{\omega}_1(t-1))$

当 $\vec{x}(t-1) = y_1(t-1)\vec{\omega}_1(t-1)$ 时，$\Delta\vec{\omega} = 0$

2.3.2. 目标函数二

$$
E(\vec{\omega}_1) = -\frac{1}{2}\ln\left(\vec{\omega}_1^\top[C_x]\vec{\omega}_1\right) + \frac{1}{2}\vec{\omega}_1^\top\vec{\omega}_1
$$

$\because \lim\limits_{\vec{\omega}_1 \to \infty} E(\vec{\omega}_1) = \infty \quad \lim\limits_{\vec{\omega}_1 \to 0} E(\vec{\omega}_1) = \infty \quad \therefore E(\vec{\omega}_1)$ 在 $(0, \infty)$ 之间存在极值

$E(\vec{\omega}_1)$ 取最小值处应满足 $\begin{cases} \nabla E(\vec{\omega}_1) = 0 \\ \nabla^2 E(\vec{\omega}_1) > 0 \end{cases}$

设最大特征向量为 $\vec{u}_{max1}$，$\nabla E(\vec{\omega}_1) = -\dfrac{[C_x]\vec{\omega}_1}{\vec{\omega}_1^\top[C_x]\vec{\omega}_1} + \vec{\omega}_1$

$$
\nabla E(\vec{\omega}_1)|_{\vec{\omega}_1=\vec{u}_{max1}} = -\frac{[C_x]\vec{u}_{max1}}{\vec{u}_{max1}^\top[C_x]\vec{u}_{max1}} + \vec{u}_{max1} = -\frac{\lambda_{max1}\vec{u}_{max1}}{\vec{u}_{max1}^\top\lambda_{max1}\vec{u}_{max1}} + \vec{u}_{max1} = 0
$$

$$
\begin{aligned}
\nabla^2 E(\vec{\omega}_1)|_{\vec{\omega}_1=\vec{u}_{max1}} &= \left.\left([\mathrm{I}] - \frac{[C_x]\vec{\omega}_1}{\vec{\omega}_1^\top[C_x]\vec{\omega}_1} + \frac{2[C_x]\vec{\omega}_1\vec{\omega}_1^\top[C_x]}{\left(\vec{\omega}_1^\top[C_x]\vec{\omega}_1\right)^2}\right)\right|_{\vec{\omega}_1=\vec{u}_{max1}} \\
&= [\mathrm{I}] - \frac{[C_x]}{\lambda_{max1}} + \frac{2\lambda_{max1}\vec{u}_{max1}(\lambda_{max1}\vec{u}_{max1})^\top}{\lambda_{max1}^2} \\
&= [\mathrm{I}] - \frac{[C_x]}{\lambda_{max1}} + 2\vec{u}_{max1}\vec{u}_{max1}^\top
\end{aligned}
$$

对于 $\vec{u}_i^\top$，$\nabla^2 E(\vec{u}_{max1})\vec{u}_i = \vec{u}_i^\top\vec{u}_i - \dfrac{\vec{u}_i^\top[C_x]\vec{u}_i}{\lambda_{max1}} + 2\vec{u}_i^\top\vec{u}_{max1}\vec{u}_{max1}^\top\vec{u}_i$

当 $\vec{u}_i = \vec{u}_{max1}$ 时，上式 $= 1 - 1 + 2 = 0$

当 $\vec{u}_i \neq \vec{u}_{max1}$ 时，$\lambda_i \leqslant \lambda_{max1}$，上式 $= 1 - \dfrac{\lambda_i}{\lambda_{max1}} > 0$

$\therefore\ \nabla^2 E(\vec{u}_{max1}) > 0$，正定

由梯度法可得 $\Delta\vec{\omega}_1 = -\eta\nabla E(\vec{\omega}_1) = -\eta\left(\dfrac{[C_x]\vec{\omega}_1}{\vec{\omega}_1^\top[C_x]\vec{\omega}_1} + \vec{\omega}_1\right)$

$$
\begin{aligned}
&= \frac{\eta}{\vec{\omega}_1^\top[C_x]\vec{\omega}_1}\left([C_x]\vec{\omega}_1 - \vec{\omega}_1^\top[C_x]\vec{\omega}_1\vec{\omega}_1\right) \\
&= \mu\left(\vec{x}\vec{x}^\top\vec{\omega}_1 - \vec{\omega}_1^\top\vec{x}\vec{x}^\top\vec{\omega}_1\vec{\omega}_1\right) \\
&= \mu\left(y_1\vec{x} - y_1^2\vec{\omega}_1\right) \\
&= \mu y_1(\vec{x} - y_1\vec{\omega}_1)
\end{aligned}
$$

2.4. 采用神经网络求解多个主元

$$\Delta\vec{\omega}_j(t) = \mu y_j\left(\vec{x} - \sum_{i\leqslant j} y_i\vec{\omega}_i\right) \quad \mu > 0$$

- 输入样本移除均值
- 根据上述公式由大到小依次求出主元

特点：不需求 $[C_x]$ 的特征向量、自动排序

2.5. Candid Covariance Free Incremental PCA (CCFI PCA)

输入样本 $\vec{x}^p$ 已移除均值，$[C_x] = [R_x] = E\left\{\vec{x}\vec{x}^\top\right\}$

$\vec{u} : [C_x]$ 的特征向量　$\lambda : [C_x]$ 的特征根　$\lambda\vec{u} = [C_x]\vec{u}$

令 $\vec{\omega}(t) = \lambda\vec{u}(t)$，则可知 $\vec{\omega}(t) = [C_x]\vec{u}(t) = \left(\dfrac{1}{t}\sum_{i=1}^{t}\vec{x}^i\vec{x}^{i\top}\right)\vec{u}(t)$

$\Rightarrow ||\vec{\omega}(t)|| = \sqrt{(\lambda\vec{u}(t))^\top(\lambda\vec{u}(t))} = \lambda \ \Rightarrow\ \vec{u}(t) = \dfrac{\vec{\omega}(t)}{\lambda} = \dfrac{\vec{\omega}(t)}{||\vec{\omega}(t)||}$

由此可知

$$\vec{\omega}(t)=\left(\frac{1}{t}\sum_{i=1}^{t}\vec{x}^{i}\vec{x}^{i\top}\right)\frac{\vec{\omega}(t-1)}{||\vec{\omega}(t-1)||}=\left(\frac{1}{t-1}\sum_{i=1}^{t-1}\vec{x}^{i}\vec{x}^{i\top}\frac{\vec{\omega}(t-1)}{||\vec{\omega}(t-1)||}\right)\frac{t-1}{t}+\frac{1}{t}\vec{x}^{t}\vec{x}^{t\top}\frac{\vec{\omega}(t-1)}{||\vec{\omega}(t-1)||}$$
$$=\frac{t-1}{t}\vec{\omega}(t-1)+\frac{1}{t}\vec{x}^{t}\vec{x}^{t\top}\frac{\vec{\omega}(t-1)}{||\vec{\omega}(t-1)||}$$

$\vec{x}^{t}$: t 时刻输入的样本（去除均值）

$$\vec{x}_1^{t}=\vec{x}^{t}\quad \vec{x}_2^{t}=\vec{x}_1^{t}-\left(\vec{x}_1^{t\top}\frac{\vec{\omega}(t)}{||\vec{\omega}(t)||}\right)\frac{\vec{\omega}(t)}{||\vec{\omega}(t)||}\quad \vec{\omega}_2(t)=\frac{t-1}{t}\vec{\omega}(t-1)+\frac{1}{t}\vec{x}_2^{t}\left(\vec{x}_2^{t}\right)^{\top}\frac{\vec{\omega}_2(t-1)}{||\vec{\omega}_2(t-1)||}$$

$$\vec{x}_i^{t}=\vec{x}_{i-1}^{t}-\left(\vec{x}_{i-1}^{t}\right)^{\top}\frac{\vec{\omega}_{i-1}(t-1)}{||\vec{\omega}_{i-1}(t-1)||}\cdot\frac{\vec{\omega}_{i-1}(t-1)}{||\vec{\omega}_{i-1}(t-1)||}$$

其中 $\vec{\omega}_{i-1}(t)=\dfrac{t-1}{t}\vec{\omega}_{i-1}(t-1)+\dfrac{1}{t}\left(\vec{x}_{i-1}^{t}\right)^{\top}\dfrac{\vec{\omega}_{i-1}(t-1)}{||\vec{\omega}_{i-1}(t-1)||}$

具体算法:

❖ *For* $t=1,2,\cdots,N$，输入样本 $\vec{x}^{t}$ 的均值 $\begin{cases}\text{增量计算 } \vec{\bar{x}}^{t}=\dfrac{t-1}{t}\vec{\bar{x}}^{t-1}+\dfrac{1}{t}\vec{x}^{t}\\ \vec{x}_1^{t}=\vec{x}^{t}-\vec{\bar{x}}^{t}\end{cases}$

  ➢ *For* $i=1,2,\cdots,\min(t,m)$, m: 降低后的维度

    ■ 若 $i=t=1,\vec{\omega}_1(1)=\vec{x}_1^{t}$

    ■ 若 $i=t\neq 1,\vec{\omega}_i(t)=\vec{x}_i^{t}$

    ■ 若 $i\neq t$ 且 $i<t$，则 $\vec{\omega}_i(t)=\dfrac{t-1}{t}\vec{\omega}_i(t-1)+\dfrac{1}{t}\vec{x}_i^{t}\left(\vec{x}_i^{t}\right)^{\top}\dfrac{\vec{\omega}_i(t-1)}{||\vec{\omega}_i(t-1)||}$

$$\vec{x}_{i+1}^{t}=\vec{x}_i^{t}-\vec{x}_i^{t}\frac{\vec{\omega}_i(t-1)}{||\vec{\omega}_i(t-1)||}\cdot\frac{\vec{\omega}_i(t-1)}{||\vec{\omega}_i(t-1)||}$$

  ➢ END

❖ END