C1-13

Machine Learning
by Andrew Ng, Stanford Engineering

Xiaojie Zhou

szxjzhou@163.com

2016.9.20

第十四集：主成分分析法

- 因子分析(Factor Analysis)
- 主成分分析(Principal Components Analysis (PCA))

第十四集：主成分分析法

- 因子分析
  - 在上一节中对于因子分析模型的优化问题进行了阐述并得到了样本x和类标z之间的联合概率分布p(x, z)，边缘概率分布p(x),p(z)和条件概率分布p(x|z),p(z|x)

$$p(z,x)=\begin{bmatrix}z\\x\end{bmatrix}\sim\mathcal{N}\left(\begin{bmatrix}\vec{0}\\\mu\end{bmatrix},\begin{bmatrix}I & \Lambda^T\\\Lambda & \Lambda\Lambda^T+\Psi\end{bmatrix}\right)$$

$$p(z)\sim\mathcal{N}\left(\vec{0},I\right)$$

$$p(x)\sim\mathcal{N}\left(\mu,\Lambda\Lambda^T+\Psi\right)$$

$$\mathrm{p(x|z)}\sim\mathcal{N}\left(\mu+\Lambda I^{-1}\left(z-\vec{0}\right),\Lambda\Lambda^T+\Psi-\Lambda I^{-1}\Lambda^T\right)=\mathcal{N}\left(\mu+\Lambda z,\Psi\right)$$

$$\mathrm{p(z|x)}\sim\mathcal{N}\left(\Lambda^T\left(\Lambda\Lambda^T+\Psi\right)^{-1}(x-\mu),I-\Lambda^T\left(\Lambda\Lambda^T+\Psi\right)^{-1}\Lambda\right)$$

  - 并由此得到了极大似然表达式

这里另外注明一点，在因子分析模型中为了简化模型，一般会给Ψ加以限制，使其成为为对角阵，即忽略Ψ的协方差成分

$$\ell(\mu,\Lambda,\Psi)=\log\prod_{i=1}^{m}\frac{1}{(2\pi)^{n/2}|\Lambda\Lambda^T+\Psi|}\exp\left(-\frac{1}{2}(x^{(i)}-\mu)^T(\Lambda\Lambda^T+\Psi)^{-1}(x^{(i)}-\mu)\right)$$

第十四集：主成分分析法

- 因子分析
  - 下面将通过EM算法对于模型的参数进行求解
    - 重复迭代下面两步直至收敛
      - 1、E-step：在已知参数θ的情况下更新每个样本x对应类标z的分布Q

$$Q_i(z^{(i)}) = \frac{1}{(2\pi)^{k/2}|\Sigma_{z^{(i)}|x^{(i)}}|^{1/2}} \exp\left(-\frac{1}{2}(z^{(i)} - \mu_{z^{(i)}|x^{(i)}})^T \Sigma^{-1}_{z^{(i)}|x^{(i)}}(z^{(i)} - \mu_{z^{(i)}|x^{(i)}})\right)$$

$$\mu_{z^{(i)}|x^{(i)}} = \Lambda^T(\Lambda\Lambda^T + \Psi)^{-1}(x^{(i)} - \mu) \quad \Sigma_{z^{(i)}|x^{(i)}} = I - \Lambda^T(\Lambda\Lambda^T + \Psi)^{-1}\Lambda$$

      - 2、M-step：在已知类标z的分布Q的情况下用极大似然法计算出参数θ

$$\sum_{i=1}^{m} \int_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \mu, \Lambda, \Psi)}{Q_i(z^{(i)})} dz^{(i)}$$

注意到在因子分析模型中类标z是服从连续值分布的，而不是离散值分布的，因此这里采用积分而非求和的方式进行计算。此处参数θ包括μ,Λ和Ψ

第十四集：主成分分析法

- 因子分析
  - 在因子分析模型中E-step的运算非常简单（在μ,Λ,Ψ这些参数均已知的情况下可以很容易求出p(z|x)），下面来介绍M-step是如何进行优化的

$$\sum_{i=1}^{m} \int_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \mu, \Lambda, \Psi)}{Q_i(z^{(i)})} dz^{(i)}$$

↓

$$\sum_{i=1}^{m} \int_{z^{(i)}} Q_i(z^{(i)}) \left[\log p(x^{(i)}|z^{(i)}; \mu, \Lambda, \Psi) + \log p(z^{(i)}) - \log Q_i(z^{(i)})\right] dz^{(i)}$$

$$= \sum_{i=1}^{m} \mathrm{E}_{z^{(i)} \sim Q_i} \left[\log p(x^{(i)}|z^{(i)}; \mu, \Lambda, \Psi) + \log p(z^{(i)}) - \log Q_i(z^{(i)})\right]$$

↓

$$\sum_{i=1}^{m} \mathrm{E} \left[\log p(x^{(i)}|z^{(i)}; \mu, \Lambda, \Psi)\right]$$

左上为M-step中需要优化的式子，由于这个式子中Q(z)的表达式和p(x|z)的表达式我们都是知道的，因此最简单也是最容易想到的办法就是将这两个表达式展开，然后进行优化。这样做确实能做出来，但是会遇到一个非常严重的问题，展开非常复杂，项数很多，给求解带来了不便。因此这时我们不妨先保留这种形式，然后考察式子中是否存在无关项，将无关项消除后再展开

通过观察左上式子发现，log的部分可以展开为log[p(x,z)]和log[Q(z)]两项相减，而其中log[p(x,z)]又可以进一步展开为log[p(x|z)]和log[p(z)]相加的形式，从而得到如左的式子。然后我们会发现，log[p(z)]和log[Q(z)]和我们要优化的变量μ,Λ,Ψ无关（p(z)分布的均值为0，方差为I，均已知；而Q(z)的值由E-step代入，同样已知），因此极大似然时可以直接忽略这些项（因为求偏微分后一定为0）

与此同时还可将形式做一个改写（这只是形式上做一个改写，其运算实质没有变），由于积分部分是对于类标z进行积分，而Q(z)是z的分布。因此积分部分相当于对于整个log函数求关于类标z的期望，由此得到一种期望表达的等价形式（再度重申，这只是表达方式不同，运算实质没有变，但这并不代表这样做完全没有意义）。将积分改写为期望后直接使得我们避开了复杂的积分运算，大大降低了后面的运算成本

由此最终得到了左下的式子，下一步就是要极大化这个式子

第十四集：主成分分析法

- 因子分析
  - 接下来可以对之前得到的式子进行展开

$$\sum_{i=1}^{m} \mathrm{E}\left[\log p(x^{(i)}|z^{(i)};\mu,\Lambda,\Psi)\right]$$

$$= \sum_{i=1}^{m} \mathrm{E}\left[\log \frac{1}{(2\pi)^{n/2}|\Psi|^{1/2}} \exp\left(-\frac{1}{2}(x^{(i)}-\mu-\Lambda z^{(i)})^T\Psi^{-1}(x^{(i)}-\mu-\Lambda z^{(i)})\right)\right]$$

$$= \sum_{i=1}^{m} \mathrm{E}\left[-\frac{1}{2}\log|\Psi| - \frac{n}{2}\log(2\pi) - \frac{1}{2}(x^{(i)}-\mu-\Lambda z^{(i)})^T\Psi^{-1}(x^{(i)}-\mu-\Lambda z^{(i)})\right]$$

  - 首先先对于参数Λ进行求解
    - 从中可以发现只有最后一项存在参数Λ，其他项不含参数Λ，对Λ求偏微分后为0，可直接忽略。因此只需考虑最后一项有

$$\nabla_{\Lambda} \sum_{i=1}^{m} -\mathrm{E}\left[\frac{1}{2}(x^{(i)}-\mu-\Lambda z^{(i)})^T\Psi^{-1}(x^{(i)}-\mu-\Lambda z^{(i)})\right]$$

第十四集：主成分分析法

- 因子分析

$$\nabla_{\Lambda}\sum_{i=1}^{m}-E\left[\frac{1}{2}\left(x^{(i)}-\mu-\Lambda z^{(i)}\right)^{T}\Psi^{-1}\left(x^{(i)}-\mu-\Lambda z^{(i)}\right)\right]=0$$

由于是对Λ求偏微分，因此不带Λ的项可直接忽略，由此可得：

$$\nabla_{\Lambda}\sum_{i=1}^{m}-E\left[-\frac{1}{2}(x^{(i)})^{T}\Psi^{-1}\Lambda z^{(i)}+\frac{1}{2}\mu^{T}\Psi^{-1}\Lambda z^{(i)}-\frac{1}{2}\left(\Lambda z^{(i)}\right)^{T}\Psi^{-1}\left(x^{(i)}-\mu-\Lambda z^{(i)}\right)\right]=0$$

由于假定Ψ为对角阵，而对角阵一定是对称阵，因此可将其中一些项进行合并，可得：

$$\nabla_{\Lambda}\sum_{i=1}^{m}-E\left[-\frac{1}{2}\left(-2\left(\Lambda z^{(i)}\right)^{T}\Psi^{-1}x^{(i)}+2\left(\Lambda z^{(i)}\right)^{T}\Psi^{-1}\mu+\left(\Lambda z^{(i)}\right)^{T}\Psi^{-1}\left(\Lambda z^{(i)}\right)\right)\right]=0$$

$$\Leftrightarrow\nabla_{\Lambda}\sum_{i=1}^{m}-E\left[-\frac{1}{2}\left(z^{(i)}\right)^{T}\Lambda^{T}\Psi^{-1}\Lambda z^{(i)}+\left(z^{(i)}\right)^{T}\Lambda^{T}\Psi^{-1}\left(x^{(i)}-\mu\right)\right]=0$$

其中发现$\left(z^{(i)}\right)^{T}\Lambda^{T}\Psi^{-1}\Lambda z^{(i)}$和$\left(z^{(i)}\right)^{T}\Lambda^{T}\Psi^{-1}\left(x^{(i)}-\mu\right)$的结果均为实数，对于实数a来说它的迹等于它本身，即$\mathrm{tr}(\mathrm{a})=\mathrm{a}$，因此原式等价为：

$$\nabla_{\Lambda}\sum_{i=1}^{m}-E\left[-tr\left(\frac{1}{2}\left(z^{(i)}\right)^{T}\Lambda^{T}\Psi^{-1}\Lambda z^{(i)}\right)+tr\left(\left(z^{(i)}\right)^{T}\Lambda^{T}\Psi^{-1}\left(x^{(i)}-\mu\right)\right)\right]=0$$

又因为矩阵的迹满足循环交换律（$\mathrm{tr}(\mathrm{ABC})=\mathrm{tr}(\mathrm{BCA})=\mathrm{tr}(\mathrm{CAB})$），因此可继续改写为：

$$\nabla_{\Lambda}\sum_{i=1}^{m}-E\left[-tr\left(\frac{1}{2}\Lambda z^{(i)}\left(z^{(i)}\right)^{T}\Lambda^{T}\Psi^{-1}\right)+tr\left(\Lambda^{T}\Psi^{-1}\left(x^{(i)}-\mu\right)\left(z^{(i)}\right)^{T}\right)\right]=0$$

又因为矩阵的迹满足$\nabla_{\mathrm{A}}tr\left(ABA^{T}C\right)=CAB+C^{T}AB$且$\nabla_{\mathrm{A}}tr\left(AB\right)=B^{T}$，因此原式等价于

$$\sum_{i=1}^{m}E\left[-\Psi^{-1}\Lambda z^{(i)}\left(z^{(i)}\right)^{T}+\Psi^{-1}\left(x^{(i)}-\mu\right)\left(z^{(i)}\right)^{T}\right]=0$$

由于这里的期望值是相对于类标$z^{(i)}$求的，因此其它量都可看成是系数被提取出来，可得：

$$\sum_{i=1}^{m}\Lambda E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]=\sum_{i=1}^{m}\left(x^{(i)}-\mu\right)E\left[\left(z^{(i)}\right)^{T}\right]$$

从而可得参数Λ的表达形式：

$$\Lambda=\left(\sum_{i=1}^{m}\left(x^{(i)}-\mu\right)E\left[\left(z^{(i)}\right)^{T}\right]\right)\left(\sum_{i=1}^{m}E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]\right)^{-1}$$

现在的问题是如何求解出这里的$\mathrm{E}\left[\left(z^{(i)}\right)^{T}\right]$和$E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]$

对于$\mathrm{E}\left[\left(z^{(i)}\right)^{T}\right]$来说相对比较简单，因为$\mathrm{E}\left[\left(z^{(i)}\right)^{T}\right]=E\left[\left(z^{(i)}\right)\right]^{T}$，而根据之前所述$Q_{i}\left(z^{(i)}\right)$表示了$z^{(i)}$的分布，因此$\mathrm{E}\left[\left(z^{(i)}\right)\right]=\mathrm{E}\left[Q_{i}\left(z^{(i)}\right)\right]=\mu_{z^{(i)}|x^{(i)}}$，故：

$$\mathrm{E}\left[\left(z^{(i)}\right)^{T}\right]=\mu_{z^{(i)}|x^{(i)}}^{T}$$

而对于$E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]$可能我们会很容易地认为$E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]=E\left[z^{(i)}\right]E\left[\left(z^{(i)}\right)^{T}\right]$，然而这是并不正确的（其实两个随机变量在相互独立时乘积的均值不一定等于这两个变量均值的乘积，只有当这两个随机变量相互独立时才能划等号，而在这里我们显然不能证明这一点）。对此正确的得到方法如下：

$$\mathrm{Var}\left(z^{(i)}\right)=E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]-E\left[z^{(i)}\right]E\left[\left(z^{(i)}\right)^{T}\right]\Leftrightarrow E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]=\mathrm{Var}\left(z^{(i)}\right)+E\left[z^{(i)}\right]E\left[\left(z^{(i)}\right)^{T}\right]$$

$$=\mathrm{Var}\left(Q_{i}\left(z^{(i)}\right)\right)+E\left[Q_{i}\left(z^{(i)}\right)\right]E\left[Q_{i}\left(z^{(i)}\right)^{T}\right]=\Sigma_{z^{(i)}|x^{(i)}}+\mu_{z^{(i)}|x^{(i)}}\mu_{z^{(i)}|x^{(i)}}^{T}$$

因此：

$$E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]=\Sigma_{z^{(i)}|x^{(i)}}+\mu_{z^{(i)}|x^{(i)}}\mu_{z^{(i)}|x^{(i)}}^{T}$$

代入$\mathrm{E}\left[\left(z^{(i)}\right)^{T}\right]$和$E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]$后可得：

$$\Lambda=\left(\sum_{i=1}^{m}\left(x^{(i)}-\mu\right)\mu_{z^{(i)}|x^{(i)}}^{T}\right)\left(\sum_{i=1}^{m}\Sigma_{z^{(i)}|x^{(i)}}+\mu_{z^{(i)}|x^{(i)}}\mu_{z^{(i)}|x^{(i)}}^{T}\right)^{-1}$$

第十四集：主成分分析法

- 因子分析

下面是对于参数μ的求解，其求解过程与参数Λ大致相同。首先还是从下面这个式子开始：

$$\nabla_{\mu}\sum_{i=1}^{m}-E\left[\frac{1}{2}\left(x^{(i)}-\mu-\Lambda z^{(i)}\right)^{T}\Psi^{-1}\left(x^{(i)}-\mu-\Lambda z^{(i)}\right)\right]=0$$

由于是对μ求偏微分，因此不带μ的项可直接忽略，由此可得：

$$\nabla_{\mu}\sum_{i=1}^{m}-E\left[-\frac{1}{2}(x^{(i)})^{T}\Psi^{-1}\mu-\frac{1}{2}\mu^{T}\Psi^{-1}\left(x^{(i)}-\mu-\Lambda z^{(i)}\right)+\frac{1}{2}\left(\Lambda z^{(i)}\right)^{T}\Psi^{-1}\mu\right]=0$$

由于假定Ψ为对角阵，而对角阵一定是对称阵，因此可将其中一些项进行合并，可得：

$$\nabla_{\mu}\sum_{i=1}^{m}-E\left[-\frac{1}{2}\left(2\mu^{T}\Psi^{-1}x^{(i)}-2\mu^{T}\Psi^{-1}\left(\Lambda z^{(i)}\right)-\mu^{T}\Psi^{-1}\mu\right)\right]=0$$

$$\Leftrightarrow\nabla_{\mu}\sum_{i=1}^{m}-E\left[\frac{1}{2}\mu^{T}\Psi^{-1}\mu+\mu^{T}\Psi^{-1}\left(\Lambda z^{(i)}-x^{(i)}\right)\right]=0$$

其中发现$\mu^{T}\Psi^{-1}\mu$和$\mu^{T}\Psi^{-1}\left(\Lambda z^{(i)}-x^{(i)}\right)$的结果均为实数，对于实数a来说它的迹等于它本身，即tr(a)=a，因此原式等价于：

$$\nabla_{\mu}\sum_{i=1}^{m}-E\left[tr\left(\frac{1}{2}\mu^{T}\Psi^{-1}\mu\right)+tr\left(\mu^{T}\Psi^{-1}\left(\Lambda z^{(i)}-x^{(i)}\right)\right)\right]=0$$

又因为矩阵的迹满足循环交换律（$\mathrm{tr}(\mathrm{ABC})=\mathrm{tr}(\mathrm{BCA})=\mathrm{tr}(\mathrm{CAB})$），因此可继续改写为：

$$\nabla_{\mu}\sum_{i=1}^{m}-E\left[tr\left(\frac{1}{2}\mu\mu^{T}\Psi^{-1}\right)+tr\left(\mu^{T}\Psi^{-1}\left(\Lambda z^{(i)}-x^{(i)}\right)\right)\right]=0$$

又因为矩阵的迹满足$\nabla_{\mathrm{A}}tr\left(ABA^{T}C\right)=CAB+C^{T}AB$且$\nabla_{\mathrm{A}}tr\left(AB\right)=B^{T}$，因此原式等价于

$$\sum_{i=1}^{m}E\left[\Psi^{-1}\mu+\Psi^{-1}\left(\Lambda z^{(i)}-x^{(i)}\right)\right]=0$$

由于这里的期望值是相对于类标$z^{(i)}$求的，因此其它量都可看成是系数被提取出来，可得：

$$\sum_{i=1}^{m}\mu=-\sum_{i=1}^{m}\left(\Lambda E\left[\left(z^{(i)}\right)^{T}\right]-x^{(i)}\right)$$

从而可得参数μ的表达形式：

$$\mu=-\frac{\sum_{i=1}^{m}\left(\Lambda E\left[\left(z^{(i)}\right)^{T}\right]-x^{(i)}\right)}{m}$$

其中由于$\Lambda=\left(\sum_{i=1}^{m}\left(x^{(i)}-\mu\right)E\left[\left(z^{(i)}\right)^{T}\right]\right)\left(\sum_{i=1}^{m}E\left[z^{(i)}\left(z^{(i)}\right)^{T}\right]\right)^{-1}$，代入后可得：

$$\mu=-\frac{\sum_{i=1}^{m}\left(\frac{\sum_{j=1}^{m}\left(x^{(j)}-\mu\right)E\left[\left(z^{(j)}\right)^{T}\right]}{\sum_{j=1}^{m}E\left[z^{(j)}\left(z^{(j)}\right)^{T}\right]}E\left[\left(z^{(i)}\right)^{T}\right]-x^{(i)}\right)}{m}$$

$$=-\frac{\sum_{i=1}^{m}\left(\sum_{j=1}^{m}\left(\left(x^{(j)}-\mu\right)E\left[\left(z^{(j)}\right)^{T}\right]\right)E\left[\left(z^{(i)}\right)^{T}\right]\right)}{m\sum_{j=1}^{m}E\left[z^{(j)}\left(z^{(j)}\right)^{T}\right]}+\frac{\left(\sum_{i=1}^{m}x^{(i)}\right)\left(\sum_{j=1}^{m}E\left[z^{(j)}\left(z^{(j)}\right)^{T}\right]\right)}{m\sum_{j=1}^{m}E\left[z^{(j)}\left(z^{(j)}\right)^{T}\right]}$$

$$\Leftrightarrow\left(\sum_{i=1}^{m}\mu\right)\left(\sum_{j=1}^{m}E\left[z^{(j)}\left(z^{(j)}\right)^{T}\right]\right)$$

$$=-\sum_{i=1}^{m}\left(\sum_{j=1}^{m}\left(x^{(j)}E\left[\left(z^{(j)}\right)^{T}\right]\right)E\left[\left(z^{(i)}\right)^{T}\right]\right)+\sum_{i=1}^{m}\left(\sum_{j=1}^{m}\left(\mu E\left[\left(z^{(j)}\right)^{T}\right]\right)E\left[\left(z^{(i)}\right)^{T}\right]\right)$$

$$+\left(\sum_{i=1}^{m}x^{(i)}\right)\left(\sum_{j=1}^{m}E\left[z^{(j)}\left(z^{(j)}\right)^{T}\right]\right)$$

$$\Leftrightarrow\left(\sum_{i=1}^{m}\mu\right)\left(\sum_{j=1}^{m}E\left[z^{(j)}\left(z^{(j)}\right)^{T}\right]\right)-\sum_{i=1}^{m}\left(\sum_{j=1}^{m}\left(\mu E\left[\left(z^{(j)}\right)^{T}\right]\right)E\left[\left(z^{(i)}\right)^{T}\right]\right)$$

$$=\left(\sum_{i=1}^{m}x^{(i)}\right)\left(\sum_{j=1}^{m}E\left[z^{(j)}\left(z^{(j)}\right)^{T}\right]\right)-\sum_{i=1}^{m}\left(\sum_{j=1}^{m}\left(x^{(j)}E\left[\left(z^{(j)}\right)^{T}\right]\right)E\left[\left(z^{(i)}\right)^{T}\right]\right)\Leftrightarrow\sum_{i=1}^{m}\mu=\sum_{i=1}^{m}x^{(i)}$$

$$\Leftrightarrow\mu=\frac{\sum_{i=1}^{m}x^{(i)}}{m}$$

第十四集：主成分分析法

- 因子分析
  - 而最后一个参数方差Ψ可以直接通过方差公式得到，由此最终可得到因子分析模型中的参数μ,Λ,Ψ的表达形式

$$\Lambda = \left(\sum_{i=1}^{m}(x^{(i)}-\mu)\mu_{z^{(i)}|x^{(i)}}^{T}\right)\left(\sum_{i=1}^{m}\mu_{z^{(i)}|x^{(i)}}\mu_{z^{(i)}|x^{(i)}}^{T}+\Sigma_{z^{(i)}|x^{(i)}}\right)^{-1} \qquad \mu = \frac{1}{m}\sum_{i=1}^{m}x^{(i)}$$

$$\Phi = \frac{1}{m}\sum_{i=1}^{m}x^{(i)}x^{(i)^{T}} - x^{(i)}\mu_{z^{(i)}|x^{(i)}}^{T}\Lambda^{T} - \Lambda\mu_{z^{(i)}|x^{(i)}}x^{(i)^{T}} + \Lambda\left(\mu_{z^{(i)}|x^{(i)}}\mu_{z^{(i)}|x^{(i)}}^{T} + \Sigma_{z^{(i)}|x^{(i)}}\right)\Lambda^{T}$$

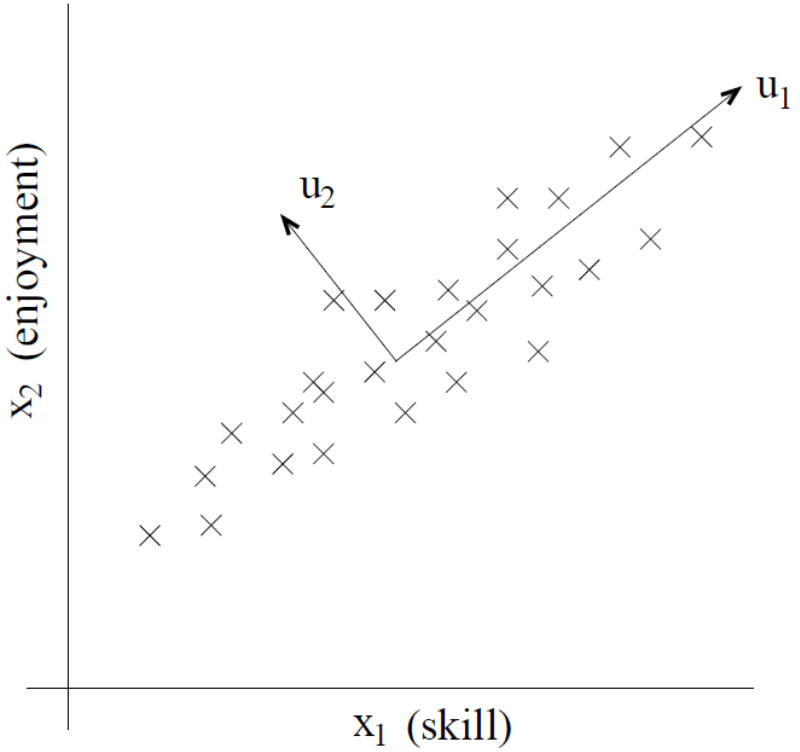
    - 不难看出这里面参数μ与$\mu_{z|x}$, $\Sigma_{z|x}$无关，因此其值不随迭代的过程而改变；同时由于要求协方差矩阵Ψ为对角阵（即忽略协方差部分），因此需要将矩阵φ中的非主对角线元素全部置为0，方可得到矩阵Ψ

第十四集：主成分分析法

- 主成分分析
  - 因子分析模型可以在一定程度上解决给定数据样本集只能生成样本空间的子空间的问题
    - 给定数据样本集只能生成样本空间的子空间的问题非常普遍，常见于下面两种情况
      - 1、样本特征多但样本数量少，导致了有些特征的信息不明显
      - 2、有些特征和其它特征本来就有很大的相关性
    - 而这样会使得高斯混合模型等的建模效果不佳或是根本无法学习出对应模型，对此解决的关键在于如何将只能生成样本空间的子空间的样本进行降维
    - 因子分析模型是一种有效的解决办法，在因子分析模型中认为高维样本x是由低维类标z通过变换$\mu+\Lambda z+\varepsilon$得到，从而间接起到降维的效果
  - 而与此同时还有一种更加直接的办法，称为主成分分析
    - 主成分分析方法比起因子分析来说更为直接，不需要经过繁琐的EM算法迭代即可达到降维的目的

第十四集：主成分分析法

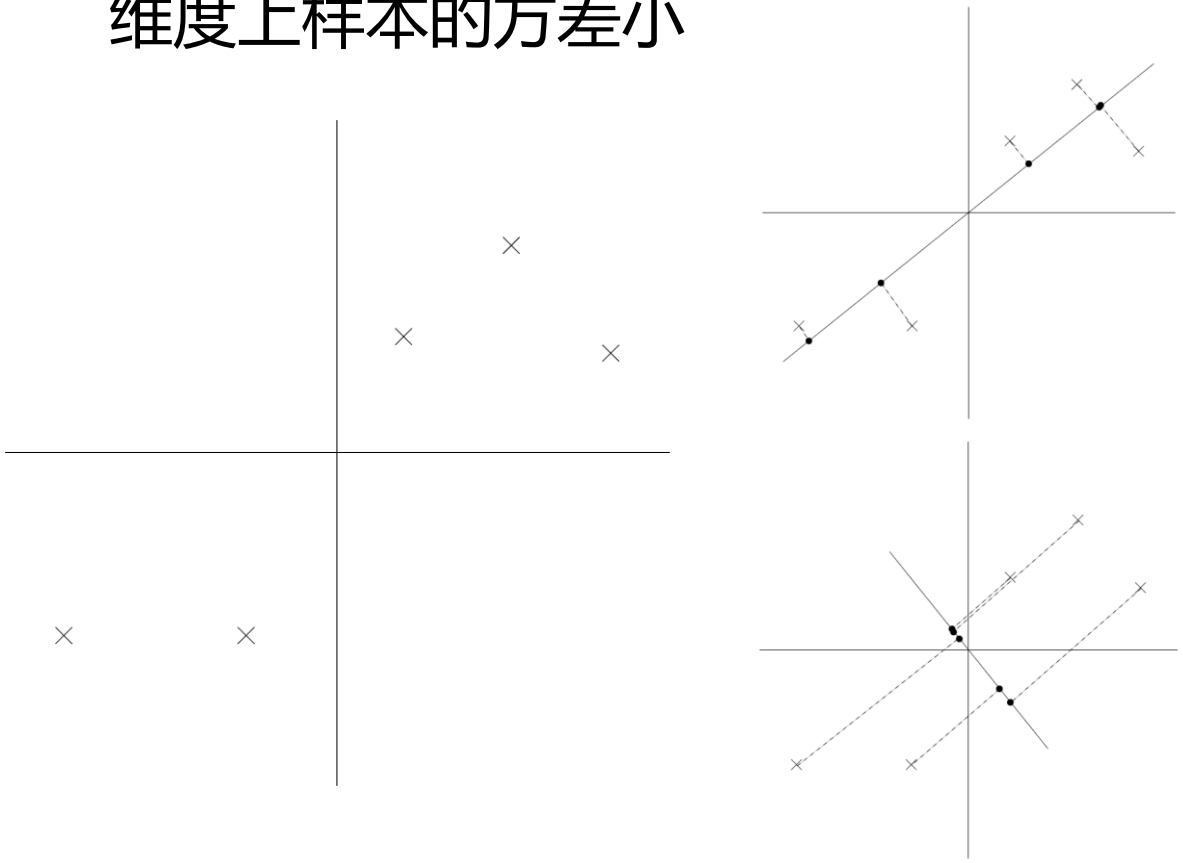
- 主成分分析
  - 在正式介绍主成分分析方法前先看一下生成样本空间的子空间的数据样本的分布特点



左边展示了驾驶员的技能和驾驶员对驾驶的兴趣这两个图线，从图中可以发现这里面数据也分布在了整个样本空间中，而非严格的子空间，但是从图中可以看出u_1方向上样本之间的区分度高（方差很大），但在u_2方向上样本之间的区分度很低（方差很小）。在这种两个特征相关性非常大情况下，即使我们已经获取了足够多的样本，但也只能近似生成样本空间的子空间，通过高斯混合模型进行学习效果不佳（u_2方向上由于方差很小，因此其值非常敏感，很容易过拟合）

第十四集：主成分分析法

- 主成分分析
  - 在上一节中揭示了数据近似生成样本空间子空间的一般性规律：在某一维度上样本的方差小



这里面更进一步地揭示了这种规律。对此得到了一种设想，对于生成样本空间的子空间的高维数据样本，我们可以在高维空间中找出一个超平面，然后将这些数据投影到这一超平面上考虑超平面上数据样本的特点，从而直接达到降维的目的

第十四集：主成分分析法

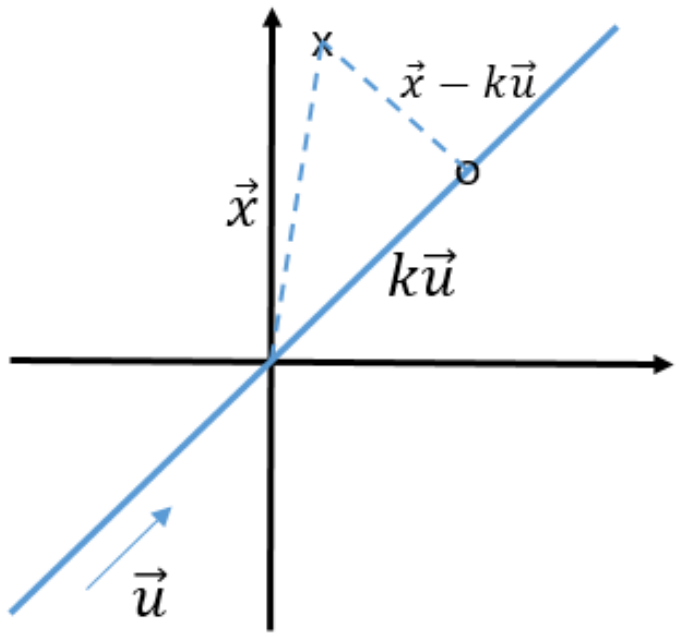
- 主成分分析
  - 那么现在的关键问题变成了我们如何找到一个这样的超平面
    - 为了更快地找到这样的超平面，我们可以先对样本做一定的预处理

1. Let $\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$.
2. Replace each $x^{(i)}$ with $x^{(i)} - \mu$.
3. Let $\sigma_j^2 = \frac{1}{m}\sum_i (x_j^{(i)})^2$
4. Replace each $x_j^{(i)}$ with $x_j^{(i)}/\sigma_j$.

从中可以看到预处理的过程就是通常意义上进行0-均值1-方差的归一化过程，在这个过程中分为四步，前面两步的作用在于去除样本的均值，从而将样本的中心移动到了原点。这样做的好处在于去除了均值这个参数，从而只用考虑过原点的超平面；而后面两步的作用在于均一化样本中每一维特征的方差，使得每一维特征的权重相同（举个例子，在这种情况下{10, 20, 30}均一化后的方差就和{1,2,3}相同），从而避免了大值的特征掩盖掉小值特征的作用

第十四集：主成分分析法

- 主成分分析
  - 下面的问题变成了如何找到使得数据投影后方差最大的那个超平面



$$\|k\vec{u}\|^2 + \|\vec{x} - k\vec{u}\|^2 = \|\vec{x}\|^2$$
$$\Leftrightarrow k^2\|\vec{u}\|^2 + \|\vec{x}\|^2 + k^2\|\vec{u}\|^2 - 2k(\vec{x})^T\vec{u} = \|\vec{x}\|^2$$
$$\Leftrightarrow 2k^2 = 2k(\vec{x})^T\vec{u} \Leftrightarrow k = (\vec{x})^T\vec{u}$$

由此可得需要最大化的目标函数：

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}(x^{(i)^T}u)^2 &= \frac{1}{m}\sum_{i=1}^{m}u^T x^{(i)}x^{(i)^T}u \\ &= u^T\left(\frac{1}{m}\sum_{i=1}^{m}x^{(i)}x^{(i)^T}\right)u\end{aligned}$$

第十四集：主成分分析法

- 主成分分析
  - 在上一步中得到了目标函数，在目标函数中不难发现下面的结论

$$u^T\left(\frac{1}{m}\sum_{i=1}^{m}x^{(i)}{x^{(i)}}^T\right)u$$

    - 1、由于之前已经去除了样本的均值，因此下式成立，可得括号部分实际上求的是样本的协方差矩阵

$$u^T\left(\frac{1}{m}\sum_{i=1}^{m}x^{(i)}\left(x^{(i)}\right)^T\right)u=u^T\left(\frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)}-E\left(x^{(i)}\right)\right)\left(x^{(i)}-E\left(x^{(i)}\right)\right)^T\right)u=u^T\Sigma u$$

    - 2、向量u为矩阵Σ的特征向量

$$\max_{u^Tu=1}u^T\Sigma u$$

这是一个仅带等式约束的优化问题，对于这个问题来说目标函数$u^T\Sigma u$中Σ已知，因此这是一个关于u的二次函数，为凸函数；但是对于等式约束$u^Tu-1=0$来也是关于u的二次函数，并不是仿射函数，因此整个问题并不是凸优化问题，因此最终结果只能保证是局部最优的，而不能保证是全局最优的。但我们可以观察一下局部最优的这些解有什么特点。因此我们先写出对应的 Lagrange 方程

$$\mathcal{L}(u,\lambda)=u^T\Sigma u+\lambda(u^Tu-1)$$

对本式中的u求偏微分即可得到所有局部最优的解u所具有的形式

$$\nabla_u\mathcal{L}(u,\lambda)=2\Sigma u+2\lambda u=0\Leftrightarrow\Sigma u=-\lambda u$$

由于Σ为矩阵，u为向量而λ为参数值，因此从这里可以得到u为矩阵Σ的特征向量（其中对应特征值最大的那个向量u称为主特征向量），$-\lambda$为矩阵Σ的特征值

第十四集：主成分分析法

- 主成分分析
  - 从之前的论述中发现，向量u为矩阵Σ的特征向量，其中矩阵Σ为样本的协方差矩阵
    - 由于u为矩阵Σ的特征向量，因此存在某个值λ使得$\Sigma u=\lambda u$. 对应到我们的目标要极大化$u^T\Sigma u$上，相当于要极大化$u^T\lambda u$
    - 由于λ为某个实数值，因此$u^T\lambda u$等价于$\lambda u^T u$，而由于u为单位向量，因此$u^T u$等于1，因此最终的目标相当于要找出矩阵Σ的最大特征值λ，而这个特征值对应的特征向量u即为能使样本投影的方差最大的一维直线方向向量
      - 矩阵最大特征值λ对应的特征向量也被称为主特征向量
      - 这里再顺便提一下特征方程$Au=\lambda u$（其中A为矩阵，u为向量，λ为实数），这可以看做是矩阵A拉伸（或压缩）了向量u，而拉伸（或压缩）的程度由参数λ所表示
        - 参数λ的正负显示了拉伸（或压缩）的方向

第十四集：主成分分析法

- 主成分分析
  - 之前讲到矩阵Σ最大特征值λ对应的主特征向量u即为能使样本投影的方差最大的一维直线方向向量
  - 现在的问题是，如果不想将样本降到一维，而是想降到任意k维（其中k的值不能大于原样本特征所生成的子空间维数）
    - 最容易想到的办法，将所有样本去掉已选择的这一维特征向量u的作用后再构造新的协方差矩阵Σ，再重新计算特征值和特征向量（相当繁琐）
    - 而实际上协方差矩阵Σ的实对称性使得我们很容易可以计算出任意k维下的主特征向量u（因为在线性代数中证明过：实对称矩阵的特征向量相互正交）
      - 因此我们只需要将满足特征方程Σu=λu的所有参数λ找出并由大到小排序，选取前k个参数λ，求出对应的特征向量u即为降维到的最优k维空间的正交基
      - 而通过这些正交基即可将原样本空间的样本降维到最优的k维空间中
        - 在这里假设原样本空间的样本x，{u1,...,uk}为得到的最优k维空间的正交基，降维后的结果为y可由下式得到：

$$y^{(i)} = \begin{bmatrix} u_1^T x^{(i)} \\ u_2^T x^{(i)} \\ \vdots \\ u_k^T x^{(i)} \end{bmatrix} \in \mathbb{R}^k$$