ML in-class exercise 10 - Dimensionality Reduction and Matrix Factorization 1

Monday, 18. January 2021 20:28

In-class Exercises

Problem 1: In this exercise, we use proof by induction to show that the linear projection onto an M-dimensional subspace that maximizes the variance of the projected data is defined by the M eigenvectors of the data covariance matrix S, given by

$$S = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^T \qquad \bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n$$

corresponding to the M largest eigenvalues. In Section 12.1 in Bishop this result was proven for the case of $M = 1$. Now suppose the result holds for some general value of M and show that it consequently holds for dimensionality $M + 1$. 1 To do this, first set the derivative of the variance of the projected data with respect to a vector $\boldsymbol{u}_{M+1}$ defining the new direction in data space equal to zero. This should be done subject to the constraints that $\boldsymbol{u}_{M+1}$ be orthogonal to the existing vectors $\boldsymbol{u}_1, \ldots \boldsymbol{u}_M$, and also that it be normalized to unit length. Use Lagrange multipliers to enforce these constraints. 2 Then make use of the orthonormality properties of the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$ to show that the new vector $\boldsymbol{u}_{M+1}$ is an eigenvector of S. 3 Finally, show that the variance is maximized if the eigenvector is chosen to be the one corresponding to eigenvector λ_{M+1} where the eigenvalues have been ordered in decreasing value.

0 Mean and variance $u_i^T x_n$

$$E[u_i^T X] = \frac{1}{N}\sum_{n=1}^{N} u_i^T x_n = u_i^T \underbrace{\left(\frac{1}{N}\sum_{n=1}^{N} x_n\right)}_{\bar{x}} = u_i^T \bar{x}$$

$$Var[u_i^T X] = \frac{1}{N}\sum_{n=1}^{N}[u_i^T(\cancel{u_i^T} x_n - \cancel{u_i^T}\bar{x})]\underbrace{[u_i^T(\cancel{u_i^T} x_n - \cancel{u_i^T}\bar{x})]^T}_{[(x_n-\bar{x})u_i]}$$

$$= u_i^T \underbrace{\left[\frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x})(x_n-\bar{x})^T\right]}_{S} u_i$$

$$= u_i^T S u_i$$

1 Construct Lagrangian $\max_{u_{M+1}} Var(u_{M+1}^T X)$

s.t. orthonormality & normalization constraints

scalar variables

$$L(u_{M+1}, \lambda_{M+1}, \eta_{1:M}) = u_{M+1}^T S u_{M+1} + \lambda_{M+1}\underbrace{(1 - u_{M+1}^T u_{M+1})}_{=0 \text{ iff } \|u_{M+1}\|=1} + \sum_{i=1}^{M}\eta_i \underbrace{u_{M+1}^T u_i}_{=0 \text{ iff } u_{M+1}\perp u_i}$$

$$\frac{\partial L}{\partial u_{M+1}} = 2Su_{M+1} - 2\lambda_{M+1}u_{M+1} + \sum_{i=1}^{M}\eta_i u_i = 0 \qquad | \cdot u_{M+1}^T \text{ from left}$$

$$2u_{M+1}^T S u_{M+1} - 2\lambda_{M+1}\underbrace{u_{M+1}^T u_{M+1}}_{1} + \sum_{i=1}^{M}\eta_i \underbrace{u_{M+1}^T u_i}_{0}$$

$$2u_{M+1}^T S u_{M+1} - 2\lambda_{M+1} = 0$$

$$\Rightarrow \underline{\underline{u_{M+1}^T S u_{M+1} = \lambda_{M+1}}} \qquad | \cdot u_{M+1} \text{ from left}$$

$$S u_{M+1} = \lambda_{M+1} u_{M+1}$$

$\Rightarrow u_{M+1}$ must be an eigenvector & λ_{M+1} it's corresponding eigenvalue

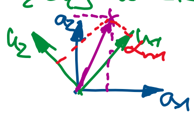
③ $\max u_{M+1}^T S u_{M+1} = \max \lambda_{M+1}$

$\Rightarrow$ to maximize choose λ_{M+1} to be the largest "remaining" eigenvalue $(= M+1^{th}$ eigenvalue$)$

Problem 2: Proof that minimizing the error is equivalent to maximizing the variance.

See Bishop chapter 12.1.2.

Preliminaries complete, orthonormal set of D-dimensional basis vectors $\{\vec u_i\}$ where $i = 1, \dots D$



$$\vec x_n = \sum_{i=1}^{D} \alpha_{ni} \vec u_i$$

wog
$$= \sum_{i=1}^{D} (\vec u_i^T \vec x_n) \vec u_i$$

still a vector

rank M projection

Residuals

we approximate
$$\vec x_n \approx \tilde x_n = \sum_{i=1}^{M} z_{ni} \vec u_i + \sum_{i=M+1}^{D} b_i \vec u_i$$

a variable for each sample n

ONE variable for the complete data

Objective:
$$\min_{\substack{\tilde X \\ \text{rank}(\tilde X) = M}} \sum_{n=1}^{N} \|\vec x_n - \tilde x\|^2$$

$$\vec x - \tilde x = \sum_{i=1}^{M} \underbrace{(\vec u_i^T \vec x_n - z_{ni})}_{=0} \vec u_i + \sum_{i=M+1}^{D} (\vec u_i^T \vec x_n - b_i) \vec u_i$$

$$\left(\frac{\partial \, \vec x - \tilde x}{\partial z_{ni}} = \dots = 0 \quad \& \quad \text{solve} \right)$$

it is obvious that $\sum_{n=1}^{N} \|\vec x_n - \tilde x\|$ is minimized if the M "first" components are 0

$$\Rightarrow \vec x - \tilde x = \sum_{i=M+1}^{D} (\vec u_i^T \vec x_n - b_i) \vec u_i$$

$$J = \sum_{n=1}^{N} \|\vec x_n - \tilde x\| = \sum_{n=1}^{N} \sum_{\substack{i=M+1 \\ i\text{-th}}}^{D} [(\vec u_i^T \vec x_n - b_i) \vec u_i]^T [(\vec u_i^T \vec x_n - b_i) \vec u_i]$$

$$\frac{\partial J}{\partial b_i} = \sum_{n=1}^{N} \cancel{2} [(\vec u^T \vec x_n - b_i) \vec u_i]^T (-\vec u_i) = 0$$

$$= \sum_{n=1}^{N} b_i \cancelto{1}{\vec u_i^T \vec u_i} - (\vec u_i^T \vec x_n) \cancelto{1}{\vec u_i^T \vec u_i}$$

$$= N b_i - \sum_{n=1}^{N} \vec u_i^T \vec x_n$$

$$\Rightarrow b_i = \frac{1}{N} \sum_{n=1}^{N} \vec u_i^T \vec x_n = \vec u^T \bar x$$

$$J = \sum_{n=1}^{N} \sum_{i=M+1}^{D} (\vec u_i^T \vec x_n - \vec u_i^T \bar x)^2 = \underline{\underline{\vec u_i^T S \vec u_i}}$$

Langrangian:

$$L(u_{M+1}, \lambda_{M+1}, \eta_{i}) = u_{M+1}^T S u_{M+1} + \lambda_{M+1} (1 - \vec u_{M+1}^T \vec u_{M+1}) + \sum \eta_i \vec u_{M+1}^T \vec u_i$$