Handed out: March 16 (Tue)

Hints for solution: Exercise class on March 19 (Fri)

Hand in: March 24 (Wed), the latest, @ Room A348

This assignment gives you maximally 5% worth of extra points for the computer assignments and the final exam.

All the exercises below have equal weight.

Ex. 1 — Gram Schmidt

1. Given two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, show that

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{a}_1 && (1) \\ \mathbf{u}_2 &= \mathbf{a}_2 - \frac{\mathbf{u}_1^T \mathbf{a}_2}{\mathbf{u}_1^T \mathbf{u}_1} \mathbf{u}_1 && (2) \end{aligned}$$

   are orthogonal to each other. Furthermore, show that any linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$ can be written in terms of $\mathbf{u}_1$ and $\mathbf{u}_2$.

2. Show by induction that for n vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$, the vectors $\mathbf{u}_k$, $k = 1, \ldots n$, are orthogonal.

$$\mathbf{u}_k = \mathbf{a}_k - \sum_{i=1}^{k-1} \frac{(\mathbf{u_i^T a_k})}{\mathbf{u}_i^T \mathbf{u}_i} \mathbf{u}_i \qquad (3)$$

Ex. 2 — Linear Algebra: Eigenvalue Decomposition

For a square matrix A of size $M \times M$, a vector $\mathbf{u}_i \neq 0$ which satisfies

$$A\mathbf{u}_i = \lambda_i \mathbf{u}_i \qquad (4)$$

is called an eigenvector of A, and λ_i is the corresponding eigenvalue. For a matrix of size $M \times M$, there are M eigenvalues λ_i (which are not necessarily distinct).

1. Show that if $\mathbf{u}_1$ and $\mathbf{u}_2$ are eigenvectors with $\lambda_1 = \lambda_2$, then $\alpha\mathbf{u}_1 + \beta\mathbf{u}_2$ is also an eigenvector with the same eigenvalue.
2. Denote by U the matrix where the column vectors are the eigenvectors $\mathbf{u}_i$ of A. Verify that the equation (4) can be written in matrix form as $AU = U\Lambda$, where Λ is a diagonal matrix with the eigenvalues λ_i as diagonal elements.

3. Show that we can write

$$\begin{aligned} A &= U\Lambda V^T, \text{ where } V^T = U^{-1} && (5)\\ A &= \sum_{i=1}^{M} \lambda_i \mathbf{u}_i \mathbf{v}_i^T && (6)\\ A^{-1} &= U\Lambda^{-1}V^T && (7)\\ A^{-1} &= \sum_{i=1}^{M} \frac{1}{\lambda_i}\mathbf{u}_i \mathbf{v}_i^T, && (8)\end{aligned}$$

Ex. 3 — Linear Algebra: Trace, Determinants and Eigenvalues

1. The trace of a matrix A is defined as $\mathrm{Tr}(A) = \sum_i a_{ii}$. Use the previous exercise to show that $\mathrm{Tr}(A) = \sum_i \lambda_i$. (You can use $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$).
2. Show that $\det A = \prod_i \lambda_i$. (Use $\det A^{-1} = 1/\det A$ and $\det(AB) = \det A \det B$ for any A and B.)

Ex. 4 — Linear Algebra: Eigenvalue Decomposition for Symmetric Matrices

1. Assume that the matrix A is symmetric, i.e. $A^T = A$. For two eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$ with $\lambda_1 \neq \lambda_2$, show that the two vectors are orthogonal to each other, that is $\mathbf{u}_1^T\mathbf{u}_2 = 0$.
2. Conclude with the results of Ex.1 that for a symmetric matrix A, all the eigenvectors $\mathbf{u}_i$ can be chosen orthogonal and of unit length (orthonormal).
3. A symmetric matrix A is said to be positive definite if $\mathbf{v}^T A\mathbf{v} > 0$ for all non-zero vectors $\mathbf{v}$. Show that positive definiteness implies that $\lambda_i > 0$, $i = 1, \ldots, M$. Show that, vice versa, $\lambda_i > 0$, $i = 1 \ldots M$ also implies that matrix A is positive definite. Conclude that positive definite matrices are invertible.

Ex. 5 — Maximum Likelihood for a Gaussian

Recall that a Gaussian random variable $X \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2 has the density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]. \tag{9}$$

1. Given iid. data $X_1, \ldots, X_N$ following a Gaussian distribution of mean μ and variance σ^2 find the likelihood $L(\mu, \sigma)$.

2. Calculate the log-likelihood $\ell(\mu, \sigma) = \log L(\mu, \sigma)$.
3. Show that the maximum likelihood estimates for the mean μ and standard deviation σ are the sample mean

$$\bar{X} = \frac{1}{N} \sum_{n=1}^{N} X_n \tag{10}$$

and the sample variance

$$S^2 = \frac{1}{N} \sum_{n=1}^{N} (X_n - \bar{X})^2. \tag{11}$$

Hint: Start from the log-likelihood. Explain why this is possible.