10907 Pattern Recognition

Lecturers
Prof. Dr. Ivan Dokmanić ⟨ivan.dokmanic@unibas.ch⟩

Tutors
Alexandra Spitzer ⟨alexandra.spitzer@unibas.ch⟩
Cheng Shi ⟨cheng.shi@unibas.ch⟩
Vinith Kishore ⟨vinith.kishore@unibas.ch⟩
Felicitas Haag ⟨felicitas.haag@unibas.ch⟩

Introductory Problem Set [Model Solutions]

This is the first weekly problem set. Its purpose is to outline the mathematical and programming background needed to navigate the course confidently. There are no weekly hand-in assignments; instead, a problem set will be posted each week and discussed in the exercise sessions. You are strongly encouraged to engage deeply with the material, as it is a prerequisite for developing the skills required to pass the exams. Remember that each exercise will not appear on the exam in exactly the same form, but it serves to outline the relevant area of knowledge. Each exercise has a difficulty level from ⋆ to ⋆⋆⋆.

1 Math

1 Linear Algebra

Exercise 1 (System of Linear Equations ⋆)

Consider the following system of linear equations,

$$\begin{aligned} 2x - 7y - 9z &= k_1 \\ 3x + 3y + 4z &= k_2 \\ 11x + 2y + 3z &= k_3 \end{aligned}$$

What are the values of k_1, k_2, k_3, for which the system above has solutions? If we set $k_1 = k_2 = k_3 = 0$, what are the possible values of x, y and z?

$$\text{Consider } \begin{cases} 2x - 7y - 9z = k_1, \\ 3x + 3y + 4z = k_2, \\ 11x + 2y + 3z = k_3. \end{cases} \qquad A = \begin{pmatrix} 2 & -7 & -9 \\ 3 & 3 & 4 \\ 11 & 2 & 3 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}.$$

Note that the rows of A are linearly dependent:

$$-(2, -7, -9) - 3(3, 3, 4) + (11, 2, 3) = \mathbf{0}$$

For the system $A\mathbf{x} = \mathbf{k}$ to be consistent, the same combination must vanish on the right-hand side:

$$-k_1 - 3k_2 + k_3 = 0 \quad \Longleftrightarrow \quad k_3 = k_1 + 3k_2.$$

Thus, *the system has solutions precisely for those (k_1, k_2, k_3) with $k_3 = k_1 + 3k_2$.*
For $k_1 = k_2 = k_3 = 0$ row-reducing A gives

$$\left[\begin{array}{ccc|c} 2 & -7 & 9 & 0 \\ 3 & 3 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{R_2 \leftarrow 2R_2 - 3R_1} \left[\begin{array}{ccc|c} 2 & -7 & 9 & 0 \\ 0 & 27 & 35 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

when choosing z as the free parameter the solutions of $A\mathbf{x} = \mathbf{0}$ satisfy

$$x = -\tfrac{1}{27} z, \qquad y = -\tfrac{35}{27} z.$$

Letting $z = 27t$ with $t \in \mathbb{R}$, we obtain the solution set

$$(x, y, z) = t\,(-1, -35, 27), \qquad t \in \mathbb{R}.$$

Exercise 2 (Elementary Row and Column Operations $\star$)

Let $\boldsymbol{M}$ be a block matrix given as

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{AB} \end{bmatrix},$$

where $\boldsymbol{A} \in \mathbb{R}^{s\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times t}$ and $\boldsymbol{I}_n$ is an $n \times n$ identity matrix. Use elementary row and column operations to transform this matrix into

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{I}_n \\ \boldsymbol{0} & \boldsymbol{A} \end{bmatrix}.$$

Provide an explanation of the row and column operations you used.

$$M = \begin{bmatrix} I_n & 0 \\ 0 & AB \end{bmatrix} \xrightarrow{C_2 \leftarrow C_2 + B\,C_1} \begin{bmatrix} I_n & B \\ 0 & AB \end{bmatrix} \xrightarrow{\text{swap }(C_1, C_2)} \begin{bmatrix} B & I_n \\ AB & 0 \end{bmatrix}$$
$$\xrightarrow{R_2 \leftarrow R_2 - A\,R_1} \begin{bmatrix} B & I_n \\ 0 & -A \end{bmatrix} \xrightarrow{R_2 \leftarrow -R_2} \begin{bmatrix} B & I_n \\ 0 & A \end{bmatrix} = N.$$

Exercise 3 (Rank of a Matrix $\star$)

Let $\boldsymbol{M}$ and $\boldsymbol{N}$ be as in the previous exercise, $\boldsymbol{M} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{AB} \end{bmatrix}$ and $\boldsymbol{N} = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{I}_n \\ \boldsymbol{0} & \boldsymbol{A} \end{bmatrix}$, with dimensions as before.

1. What conclusions can you draw regarding the rank of matrix $\boldsymbol{N}$ from the ranks of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$?
2. Using the results from part 1 of the exercise and the previous exercise, prove theinequality: $\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}) - \operatorname{rank}(\boldsymbol{AB}) \leq n$.
3. If $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{AB}) = s$, what is the rank of matrix $\boldsymbol{B}$?

1. The rank of a matrix is invariant und elementary row and column operations $\Longrightarrow$ $\operatorname{rank}(\boldsymbol{M}) = \operatorname{rank}(\boldsymbol{N})$. Since $\boldsymbol{M}$ is a block diagnoal matrice, the rank can be expressed simply as $\operatorname{rank}(\boldsymbol{N}) = \operatorname{rank}(\boldsymbol{I_n}) + \operatorname{rank}(\boldsymbol{AB}) = n + \operatorname{rank}(\boldsymbol{AB})$. Consequently,
$$n \leq \operatorname{rank}(\boldsymbol{N}) = n + \operatorname{rank}(\boldsymbol{AB}) \leq n + \operatorname{rank}(\boldsymbol{A}),$$
and equality with $n + \operatorname{rank}(\boldsymbol{A})$ holds only when $\operatorname{rank}(\boldsymbol{AB}) = \operatorname{rank}(\boldsymbol{A})$.
2. This follows from (1.) and $\operatorname{rank}(\boldsymbol{N}) \leq \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B})$.
3. If $\operatorname{rank}(\boldsymbol{B})$ is less than the number of comlumns in $\boldsymbol{A}$, then not all columns of $\boldsymbol{A}$ can be used indepedently in the linear combinations which limits the rank of $\boldsymbol{AB}$. On the other hand, if $\operatorname{rank}(\boldsymbol{A})$ is less than the number of rows in $\boldsymbol{B}$, then the linear dependence among the columns of $\boldsymbol{A}$ will carry over to $\boldsymbol{AB}$, again limiting its rank.

Exercise 4 (Eigenvalues and Positive Definiteness)

Let $\boldsymbol{A} \in \mathbb{R}^{s \times s}$ be an $s \times s$ real matrix such that $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^s$. Show that $\boldsymbol{A}$ does not have any negative eigenvalues.

Given the Eigenwert-equation $Av = \lambda v$, $v \in \mathbb{R}^s$ we get

$$\begin{aligned} Av &= \lambda v \\ v^T Av &= \lambda v^T v \\ \frac{v^T Av}{v^T v} &= \lambda. \end{aligned}$$

with $v^T Av \geq 0$ and $v^T v > 0$ as eigenvectors are not-zero by definition.

Another way to prove this: Suppose, to the contrary, that $\lambda < 0$ is an eigenvalue of A. Since λ is real, there exists a nonzero real eigenvector $v \in \mathbb{R}^s$ with $Av = \lambda v$. Then

$$v^T Av = v^T(\lambda v) = \lambda\, v^T v.$$

By hypothesis $v^T Av \geq 0$, while $v^T v > 0$ and $\lambda < 0$, so the right-hand side is negative. This is a contradiction. Hence A has no negative eigenvalues.

2 Probability

Exercise 5 (System of Linear Equations with Random Variables ⋆⋆)

For the system of linear equations given below:

$$\begin{aligned} 2X - 7Y + 9Z &= K_1 \\ 3X + 3Y + 4Z &= K_2 \\ 5X + 2Y + 5Z &= K_3 \end{aligned}$$

The random variables X, Y, Z are independent and normally distributed such that $X \sim \mathcal{N}(0,1), Y \sim \mathcal{N}(1,4)$ and $Z \sim \mathcal{N}(0,9)$.

- What are the distributions of K_1, K_2 and K_3?
- What is the joint probability density function (pdf) of K_1, K_2 and K_3, $p(k_1, k_2, k_3)$?

Recall that a normally distributed random variable X with mean μ and variance σ^2 (written $X \sim \mathcal{N}(\mu, \sigma^2)$) has the probability density function (pdf)

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

so that for some a, b such that $a \leq b$ we have

$$\mathbb{P}[a \leq X \leq b] = \int_a^b p_X(x)dx.$$

1. **Marginal distributions of K_1, K_2, K_3.**

   Write $\boldsymbol{K} = (K_1, K_2, K_3)^T$ and $\boldsymbol{U} = (X, Y, Z)^T$. Then $\boldsymbol{K} = M\boldsymbol{U}$ with

$$M = \begin{bmatrix} 2 & -7 & 9 \\ 3 & 3 & 4 \\ 5 & 2 & 5 \end{bmatrix}, \qquad \boldsymbol{\mu} = \mathbb{E}[\boldsymbol{U}] = \begin{bmatrix} \mathbb{E}[\boldsymbol{X}] \\ \mathbb{E}[\boldsymbol{Y}] \\ \mathbb{E}[\boldsymbol{Z}] \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\Sigma = \mathrm{Var}(\boldsymbol{U}) = \begin{bmatrix} \mathrm{Var}(\boldsymbol{X}) & 0 & 0 \\ 0 & \mathrm{Var}(\boldsymbol{Y}) & 0 \\ 0 & 0 & \mathrm{Var}(\boldsymbol{Z}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

Remember that $\mathbb{E}[a\boldsymbol{X} + b] = a\mathbb{E}[\boldsymbol{X}] + b$, $\mathrm{Var}(a\boldsymbol{X} + b) = a^2\,\mathrm{Var}(\boldsymbol{X})$ and for independent random variables $\mathbb{E}(\boldsymbol{X}) + \mathbb{E}(\boldsymbol{Y}) = \mathbb{E}(\boldsymbol{X} + \boldsymbol{Y})$ and $\mathrm{Var}(\boldsymbol{X}) + \mathrm{Var}(\boldsymbol{Y}) = \mathrm{Var}(\boldsymbol{X} + \boldsymbol{Y})$.

Any affine transform of a (jointly) normal vector is normal,

$$\boldsymbol{K} \sim \mathcal{N}\big(M\boldsymbol{\mu},\ M\Sigma M^T\big).$$

Hence each K_i is normal with mean equal to the corresponding row of $M\boldsymbol{\mu}$ and variance equal to the corresponding diagonal entry of $M\Sigma M^T$:

$$\begin{aligned} \mathbb{E}[K_1] &= 2\cdot 0 - 7\cdot 1 + 9\cdot 0 = -7, & \mathrm{Var}(K_1) &= 2^2\cdot 1 + (-7)^2\cdot 4 + 9^2\cdot 9 = 929; \\ \mathbb{E}[K_2] &= 3\cdot 0 + 3\cdot 1 + 4\cdot 0 = 3, & \mathrm{Var}(K_2) &= 3^2\cdot 1 + 3^2\cdot 4 + 4^2\cdot 9 = 189; \\ \mathbb{E}[K_3] &= 5\cdot 0 + 2\cdot 1 + 5\cdot 0 = 2, & \mathrm{Var}(K_3) &= 5^2\cdot 1 + 2^2\cdot 4 + 5^2\cdot 9 = 266. \end{aligned}$$

Therefore

$$K_1 \sim \mathcal{N}(-7, 929), \qquad K_2 \sim \mathcal{N}(3, 189), \qquad K_3 \sim \mathcal{N}(2, 266).$$

2. **Joint pdf of (K_1, K_2, K_3).** 这里不会算

The covariance is

$$\mathrm{Cov}(X, Y) \;=\; \mathbb{E}\big[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\big] \;=\; \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

The covariance matrix is given by

$$\Sigma_K = \mathrm{Cov}\begin{pmatrix} K_1 \\ K_2 \\ K_3 \end{pmatrix} = \begin{bmatrix} \mathrm{Var}(K_1) & \mathrm{Cov}(K_1, K_2) & \mathrm{Cov}(K_1, K_3) \\ \mathrm{Cov}(K_2, K_1) & \mathrm{Var}(K_2) & \mathrm{Cov}(K_2, K_3) \\ \mathrm{Cov}(K_3, K_1) & \mathrm{Cov}(K_3, K_2) & \mathrm{Var}(K_3) \end{bmatrix}$$

By using the formula $\mathrm{Cov}(aX+bY+cZ, dX+eY+fZ) \;= ad\,\mathrm{Var}(X)+be\,\mathrm{Var}(Y)+cf\,\mathrm{Var}(Z)$ we can calculate each $\mathrm{Cov}(K_i, K_j) = \sum_{\ell=1}^{3} m_{i\ell} m_{j\ell}\,\mathrm{Var}(U_\ell)$. For example $\mathrm{Cov}(K_1, K_2) = 2\cdot 3\cdot 1 + (-7)\cdot 3\cdot 4 + 9\cdot 4\cdot 9 = 246.)$

The mean vector and covariance matrix are

$$\boldsymbol{\mu}_K = M\boldsymbol{\mu} = \begin{bmatrix} -7 \\ 3 \\ 2 \end{bmatrix}, \qquad \Sigma_K = M\Sigma M^T = \begin{bmatrix} 929 & 246 & 359 \\ 246 & 189 & 219 \\ 359 & 219 & 266 \end{bmatrix}.$$

Since $\det(\Sigma_K) = 374{,}544$ and $\sqrt{\det(\Sigma_K)} = 612$, and

$$\Sigma_K^{-1} = \frac{1}{41616}\begin{bmatrix} 257 & 1465 & -1553 \\ 1465 & 13137 & -12793 \\ -1553 & -12793 & 12785 \end{bmatrix},$$

the joint pdf of $\boldsymbol{K} = (k_1, k_2, k_3)^T$ is the $\mathcal{N}_3(\boldsymbol{\mu}_K, \Sigma_K)$ density:

$$p(k_1, k_2, k_3) = \frac{1}{(2\pi)^{3/2}\,612}\,\exp\!\big(-\tfrac{1}{2}\,(\boldsymbol{k} - \boldsymbol{\mu}_K)^T \Sigma_K^{-1} (\boldsymbol{k} - \boldsymbol{\mu}_K)\big)\,.$$

Exercise 6 (Sum of Bernoulli Distribution $\star\star\star$)

Alice and Bob decide to play a game. They toss a biased coin n times. If it comes up heads at least $n/2$ times, Bob wins; otherwise Alice wins. The total number of heads is

$$S = \sum_{i=1}^{n} X_i$$

where X_i is an independent Bernoulli random variable with mean p, that is,

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1 - p.$$

Assume that $p = 0.52$.

1. What is the law of the random variable S? Give its name, its probability mass function (p.m.f.) and its mean.

2. Bob only wants to play if he has a 90% chance of winning.
   For a given n, what is the expected number of heads?
   Compute the probability that Bob wins when $n = 2$.
   Can you figure out for which values of n Bob will participate in the game?

   Hint: One way to do this is by using Hoeffding's inequality: let Z_i be independent and identically distributed random variables such that $0 \leq Z_i \leq 1$, then

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - \mathbb{E}(Z_i) \geq t\right) \leq \exp(-nt^2).$$

3. Alice and Bob continue the game for a long, long time. What can you say about the value or distribution of S as $n \to \infty$? What about the value or distribution of $\frac{S}{n}$? And what about the value or distribution of $\frac{S - np}{\sqrt{n}}$?

1. Since S is a sum of independent Bernoulli(p) random variables's,

$$S \sim \text{Binomial}(n, p), \qquad \mathbb{P}(S = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

   Hence

$$\mathbb{E}[S] = np = 0.52\,n, \qquad \text{Var}(S) = np(1-p) = n \cdot 0.52 \cdot 0.48 = 0.2496\,n.$$

   However,as $n \to \infty$, the probability mass function of S approaches a normal distribution by the central limit theorem and $S \sim \mathcal{N}(np, (1-p))$.

2. For given n, Bob's win probability is

$$\mathbb{P}\big(S \geq \lceil n/2 \rceil\big) = \sum_{k=\lceil n/2 \rceil}^{n} \binom{n}{k} p^k (1-p)^{n-k}.$$

   Expected number of heads: $\mathbb{E}[S] = 0.52\,n$.

   Example $n = 2$*:*

$$\mathbb{P}(\text{Bob wins}) = \mathbb{P}(S \geq 1) = 1 - \mathbb{P}(S = 0) = 1 - (1-p)^2 = 1 - 0.48^2 = 0.7696.$$

   Thus for $n = 2$ Bob's win chance is $76.96\% < 90\%$.

For which n would Bob play? (Hoeffding bound). Note that Bob loses iff $S/n < 1/2$. By Hoeffding's inequality for $X_i \in [0,1]$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p \le -t\right) \le e^{-nt^2} \qquad (t>0).$$

With $t = p - \frac{1}{2} = 0.02$,

$$\mathbb{P}\left(\frac{S}{n} < \frac{1}{2}\right) \le e^{-n(0.02)^2} = e^{-0.0004\,n}.$$

Requiring this to be at most 0.1 (so Bob wins with probability ≥ 0.9) gives

$$e^{-0.0004\,n} \le 0.1 \iff n \ge \frac{\ln 10}{0.0004} \approx 5756.5.$$

Hence the bound guarantees Bob's condition for all $n \ge 5757$.

3.
   - Since $\mathbb{E}[S] = 0.52\,n \to \infty$, we have $S \to \infty$ a.s. and grows linearly in n.
   - By the (strong) law of large numbers,
   $$\frac{S}{n} \to p = 0.52.$$
   - By the central limit theorem,
   $$\frac{S-np}{\sqrt{n\,p(1-p)}} \to \mathcal{N}(0,1), \qquad \text{equivalently} \qquad \frac{S-np}{\sqrt{n}} \to \mathcal{N}\big(0,\ p(1-p)\big),$$
   .

Exercise 7 (Expectations ⋆⋆)

The expectation of a non-negative random variable X with probability density function $f_X(x)$ is

$$\mathbb{E}(X) = \int_0^\infty x f_X(x)dx.$$

Show that an alternative way to compute it is by the following useful identity:

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X>t)dt.$$

Given

$$\mathbb{P}(X>t) = 1 - \mathbb{P}(X \le t) = 1 - \int_0^t f_x(x)\,dx.$$

and

$$\int_0^\infty f_x(x)\,dx = 1,$$

we can plug in the above expression such that

$$\mathbb{P}(X>t) = \int_0^\infty f_x(x)\,dx - \int_0^t f_x(x)\,dx = \int_t^\infty f_x(x)\,dx.$$

We can express the expectation as follows

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X>t)dt = \int_0^\infty \int_t^\infty f_x(x)\,dxdt = \int_0^\infty \int_0^\infty \mathbf{1}_{\{x\ge t\}}\, f_X(x)\,dx\,dt.$$

Since the integrand $\mathbf{1}_{\{x\ge t\}} f_X(x) \ge 0$, Fubini's theorem allows us to switch the order (Keep in mind that $\int_0^\infty \mathbf{1}_{\{x\ge t\}}\,dt = \int_0^x 1\,dt + \int_x^\infty 0\,dt = \big[t\big]_0^x + 0 = x$).

$$\int_0^\infty \int_0^\infty \mathbf{1}_{\{x\ge t\}}\, f_X(x)\,dt\,dx = \int_0^\infty \Big(\int_0^x 1\ dt\Big) f_X(x)\,dx = \int_0^\infty x\, f_X(x)\,dx = \mathbb{E}[X].$$

2 Coding

To get your hands dirty with Python, you are asked to write some basic array operations realted to linear algebra and probability. Specifically, you need to complete the blank functions in `sample.py`.

Exercise 8 (Basic Computation ⋆)

1. Write a function `check_span` that takes as the inputs a vector $\boldsymbol{y}$ and a matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots \boldsymbol{a}_n]$, where $\boldsymbol{a}_i$'s are the column vectors of the matrix. The function outputs a Boolean variable. The output is `True` if $\boldsymbol{y}$ is in the span of column vectors in $\boldsymbol{A}$. Otherwise, the output is `False`.

2. Write a function `estimate_mean_and_covariance` to compute the mean and the unbiased covariance matrix of a set of vectors $\{\boldsymbol{x}_1, \ldots \boldsymbol{x}_n\}$. The set of vectors are passed to the function as a matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots \boldsymbol{x}_n]$. The unbiased covariance matrix is defined as:

$$\boldsymbol{Q} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^{\mathrm{T}}$$

where $\bar{\boldsymbol{x}} = \frac{1}{n} \sum_i^n \boldsymbol{x}$.