Lesson 1: Preliminaries

Theme: Review of some basic concepts.

Linear algebra. Let \( \bar{u}_1, \ldots, \bar{u}_n \in \mathbb{R}^m \) be vectors in \( \mathbb{R}^m \). They are linearly dependent, if there are \( \alpha_1, \ldots, \alpha_n \), not all of them are zero, such that
\[
\alpha_1 \bar{u}_1 + \cdots + \alpha_n \bar{u}_n = 0
\]
They are called linearly independent, if there are no such \( \alpha_1, \ldots, \alpha_n \).

The spanning space of a set of vectors \( \bar{u}_1, \ldots, \bar{u}_n \) is the vector space:
\[
\text{span}(\bar{u}_1, \ldots, \bar{u}_n) := \{ \alpha_1 \bar{u}_1 + \alpha_2 \bar{u}_2 + \cdots + \alpha_n \bar{u}_n \mid \alpha_1, \ldots, \alpha_n \in \mathbb{R} \}.
\]
The dimension of a vector space \( V \) is:
\[
\dim(V) := \max\{ k \mid \text{there are } k \text{ linearly independent vectors in } V \}
\]
Let \( A \) be the following \((m \times m)\)-matrix over \( \mathbb{R} \):
\[
A = \begin{bmatrix}
a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,m} \\
a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,m}
\end{bmatrix}
\]
The column vectors of \( A \) are the vectors:
\[
\begin{bmatrix}
a_{1,1} \\
a_{2,1} \\
\vdots \\
a_{m,1}
\end{bmatrix}, \quad
\begin{bmatrix}
a_{1,2} \\
a_{2,2} \\
\vdots \\
a_{m,2}
\end{bmatrix}, \quad \ldots, \quad
\begin{bmatrix}
a_{1,m} \\
a_{2,m} \\
\vdots \\
a_{m,m}
\end{bmatrix}
\]
The row vectors of \( A \) are the vectors:
\[
\begin{bmatrix}
a_{1,1} \\
a_{1,2} \\
\vdots \\
a_{1,m}
\end{bmatrix}, \quad
\begin{bmatrix}
a_{2,1} \\
a_{2,2} \\
\vdots \\
a_{2,m}
\end{bmatrix}, \quad \ldots, \quad
\begin{bmatrix}
a_{m,1} \\
a_{m,2} \\
\vdots \\
a_{m,m}
\end{bmatrix}
\]
The column rank of \( A \) is the dimension of the span of its column vectors, while the row rank of \( A \) is the dimension of the span of its row vectors. It is known that (for square matrices) column rank equals row rank, and we will denote it by \( \text{rank}(A) \).

Probability space. A probability space is a system \( (\Omega, \text{Pr}) \), where \( \Omega \) is a set called sample space, and \( \text{Pr} : 2^\Omega \rightarrow \mathbb{R} \) is a probability function satisfying the following conditions.

- \( \text{Pr}[\Omega] = 1 \),
- \( 0 \leq \text{Pr}[E] \leq 1 \), for every \( E \in 2^\Omega \),
for any countably infinite sequence of pairwise disjoint sets \( E_1, E_2, \ldots \),

\[
\Pr\left[ \bigcup_{i \geq 1} E_i \right] = \sum_{i \geq 1} \Pr[E_i]
\]

The sets in \( 2^\Omega \) are usually called events, and the singletons \( \{ e \} \) elementary events.

We will only deal with discrete probability space, i.e., when \( \Omega \) is a countable set. Without loss of generality, we can assume that \( \Pr[e] > 0 \), for every \( e \in \Omega \).

We say that two events \( E \) and \( F \) are independent, if \( \Pr[E \cap F] = \Pr[E] \cdot \Pr[F] \). Likewise, a collection of events \( E_1, \ldots, E_k \) are independent, if for every \( I \subseteq \{1, \ldots, k\} \),

\[
\Pr\left[ \bigcap_{i \in I} E_i \right] = \prod_{i \in I} \Pr[E_i]
\]

The conditional probability that event \( E \) occurs given that event \( F \) occurs is defined as:

\[
\Pr[E \mid F] := \frac{\Pr[E \cap F]}{\Pr[F]}
\]

A (discrete) random variable is a function \( X : \Omega \to \mathbb{R} \). A random variable \( X \) is called a 0-1 random variable, if \( \text{range}(X) = \{0, 1\} \).

The probability of the event “\( X = a \)” is defined as:

\[
\Pr[X = a] := \sum_{e \in \Omega \text{ such that } X(e) = a} \Pr[e]
\]

The probabilities \( \Pr[X \oplus a] \), where \( \oplus \in \{\leq, \geq, <, >, \neq\} \) can be defined in a similar manner. We say that a random variable \( X \) is uniformly distributed on \( \text{range}(X) \), if \( \Pr[X = a] = \Pr[X = b] \), for every \( a, b \in \text{range}(X) \).

Two random variables \( X, Y \) are independent, if \( \Pr[X = x \cap Y = y] = \Pr[X = x] \cdot \Pr[Y = y] \), for every possible values \( x \) and \( y \). Likewise, a collection of random variables \( X_1, \ldots, X_k \) are independent, if for every \( I \subseteq \{1, \ldots, k\} \), for every \( i \in I \), for every value \( x_i \),

\[
\Pr\left[ \bigcap_{i \in I} X_i = x_i \right] = \prod_{i \in I} \Pr[X_i = x_i]
\]

The expectation of a random variable \( X \) is defined as \( \mathsf{E}[X] := \sum_i i \cdot \Pr[X = i] \). It is known that for every two random variables \( X \) and \( Y \), and for every constant \( c \),

- \( \mathsf{E}[X + Y] = \mathsf{E}[X] + \mathsf{E}[Y] \),
- \( \mathsf{E}[cX] = c\mathsf{E}[X] \).

A pair \((X, Y)\) of random variables can be viewed as a random variable \( Z \) with an appropriate “pairing function” \( \langle \cdot, \cdot \rangle \), where \( \Pr[Z = c] = \Pr[\langle X = a, Y = b \rangle] \), where \( \langle a, b \rangle = c \). For convenience, we will simply write \((X, Y)\) to denote a random variable \( Z \) obtained in this way.

Below we have some useful bounds that we will often use in this course.

**Theorem 1.1 (Markov’s inequality)** Let \( X \) be a non-negative random variables. Then,

\[
\Pr[X \geq \alpha] \leq \frac{\mathsf{E}[X]}{\alpha}
\]

**Theorem 1.2 (Chernoff’s inequality)** Let \( X_1, \ldots, X_n \) be independent 0-1 random variables with \( \Pr[X_i = 1] = p \). Let \( X = X_1 + \cdots + X_n \) and let \( \mu = \mathsf{E}[X] \). Then, for every \( 0 < \delta < 1 \),

\[
\Pr[|X - \mu| \geq \delta \mu] \leq 2e^{-\mu \delta^2/3}.
\]