# Selected Topics in Mathematics of Learning

**High-Dimensional Statistics**

Lecturer: Marius Yamakou

Winter Semester 2024/25

Department of Data Science, FAU

October 29, 2024

## Part I

### Why high-dimensional statistics?

**Objectives:**

- **Understand the importance of high-dimensional statistics:**
  - Explain why high-dimensional data is critical in modern data science and machine learning.

## Part I

### Why high-dimensional statistics?

**Objectives:**

- **Understand the importance of high-dimensional statistics:**
  - Explain why high-dimensional data is critical in modern data science and machine learning.
- **Identify common challenges in high dimensions:**
  - Discuss computational and statistical difficulties in high-dimensional data.

## Part I

### Why high-dimensional statistics?

**Objectives:**

- **Understand the importance of high-dimensional statistics:**
  - Explain why high-dimensional data is critical in modern data science and machine learning.
- **Identify common challenges in high dimensions:**
  - Discuss computational and statistical difficulties in high-dimensional data.
- **Explore specific examples**
- **Discuss solutions:**
  - Present methods like dimension reduction to handle challenges.

# Outline

# 1. Why High-Dimensional Statistics?

## High-dimensional data: Motivation

- **Intensive data collection with increasing number of features measured per individual.**
    - Biotech data (e.g., genetics: millions of genes and combinations per observation).
    - High-resolution imaging (millions of pixels/voxels).
    - Finance (e.g., stock indices).
    - Climate studies.
    - Web data.
    - Crowdsourcing data, etc

# 1.1 High-Dimensional Data: Blessing or Curse?

## Blessing

- We can sense thousands of variables on each "individual": potentially we will be able to scan every variable that may influence the phenomenon under study.

# 1.1 High-Dimensional Data: Blessing or Curse?

## Blessing

- We can sense thousands of variables on each "individual": potentially we will be able to scan every variable that may influence the phenomenon under study.

## Curse

- Separating the signal from the noise is in general almost impossible in high-dimensional data, computations can rapidly exceed the available resources, volumes can unexpectedly vanish.

# 1.2 Classical vs. High-Dimensional Statistical Theories

## 1. Classical Theory

- Assumption: $N \gg p$ (number of observations $N$ is much larger than the number of features $p$).
- Asymptotic assumption: $p$ is fixed while $N \to \infty$.
- Key tools:
  - Law of Large Numbers (LLN)
  - Central Limit Theorem (CLT)

# 1.2 Classical vs. High-Dimensional Statistical Theories

## 1. Classical Theory

- Assumption: $N \gg p$ (number of observations $N$ is much larger than the number of features $p$).
- Asymptotic assumption: $p$ is fixed while $N \to \infty$.
- Key tools:
  - Law of Large Numbers (LLN)
  - Central Limit Theorem (CLT)

## 2. High-Dimensional Theory

- Assumptions:
  - $N \sim p$, e.g., $\frac{N}{p} \to \alpha$ (finite ratio as $N, p \to \infty$).
  - $p \gg N$, e.g., $p \sim e^N$ (exponential growth of $p$ relative to $N$).
- Often non-asymptotic (finite-sample) analysis such as concentration inequalities.

# 1.2 Classical vs. High-Dimensional Statistical Theories

## 1. Classical Theory

- Assumption: $N \gg p$ (number of observations $N$ is much larger than the number of features $p$).
- Asymptotic assumption: $p$ is fixed while $N \to \infty$.
- Key tools:
  - Law of Large Numbers (LLN)
  - Central Limit Theorem (CLT)

## 2. High-Dimensional Theory

- Assumptions:
  - $N \sim p$, e.g., $\frac{N}{p} \to \alpha$ (finite ratio as $N, p \to \infty$).
  - $p \gg N$, e.g., $p \sim e^N$ (exponential growth of $p$ relative to $N$).
- Often non-asymptotic (finite-sample) analysis such as concentration inequalities.

## 3. Challenges in High Dimensions

- Number of features $p$ may exceed the number of observations $N$. E.g., 104 genes per only 50 samples.
- Not all features are relevant for answering a specific question.
- Classical methods often fail (e.g., linear regression, covariance estimation) due to the "curse of dimensionality."

## 2. What can go wrong in high dimensions?
### High-dimensional ball and volume concentration

- In high-dimensional spaces, the volume of a ball is concentrated near its surface.
- Let $B_p(0, r)$ represent an Euclidean ball of radius $r$ in $p$-dimensions.
- The volume of such a ball is: $V_p(r) = r^p V_p(1)$ where $V_p(1)$ is the volume of a unit ball in $p$-dimensions.

# 2. What can go wrong in high dimensions?
## High-dimensional ball and volume concentration

- In high-dimensional spaces, the volume of a ball is concentrated near its surface.
- Let $B_p(0, r)$ represent an Euclidean ball of radius $r$ in $p$-dimensions.
- The volume of such a ball is: $V_p(r) = r^p V_p(1)$ where $V_p(1)$ is the volume of a unit ball in $p$-dimensions.

### Crust of a High-Dimensional Ball

- Define the “crust" as the thin outer layer: $C_p(r) = B_p(0, r) \setminus B_p(0, 0.99r)$
- The fraction of the volume in the crust is: $\frac{\text{volume}(C_p(r))}{\text{volume}(B_p(0,r))} = 1 - 0.99^p$

- As $p \to \infty$, $1 - 0.99^p$ approaches 1, meaning almost all the volume is concentrated in the crust. Thus, a ball is essentially a sphere in high dimensions.

## 2. What can go wrong in high dimensions?
### High-dimensional ball and volume concentration

- In high-dimensional spaces, the volume of a ball is concentrated near its surface.
- Let $B_p(0, r)$ represent an Euclidean ball of radius $r$ in $p$-dimensions.
- The volume of such a ball is: $V_p(r) = r^p V_p(1)$ where $V_p(1)$ is the volume of a unit ball in $p$-dimensions.

**Crust of a High-Dimensional Ball**

- Define the "crust" as the thin outer layer: $C_p(r) = B_p(0, r) \setminus B_p(0, 0.99r)$
- The fraction of the volume in the crust is: $\frac{\text{volume}(C_p(r))}{\text{volume}(B_p(0,r))} = 1 - 0.99^p$

- As $p \to \infty$, $1 - 0.99^p$ approaches 1, meaning almost all the volume is concentrated in the crust. Thus, a ball is essentially a sphere in high dimensions.

**Takeaway**

In high dimensions, most of the volume is located near the surface, which defies low-dimensional intuition. This result shows that our usual intuition about shapes doesn't hold in high dimensions, which has significant implications for understanding high-dimensional data and probability distributions.

1. **Volume of a unit ball in $\mathbb{R}^p$:** $V_p(1) := \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}+1\right)}$.

   where $\Gamma$ is the gamma function, a generalization of the factorial function such that $\Gamma(p) = (p-1)! \quad$ for positive integers $p$.

1. **Volume of a unit ball in $\mathbb{R}^p$:** $V_p(1) := \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}+1\right)}$.

   where $\Gamma$ is the gamma function, a generalization of the factorial function such that $\Gamma(p) = (p-1)! \quad$ for positive integers $p$.

2. **By Stirling's approximation for large $p$:** That is $p! \approx \sqrt{2\pi p}\left(\frac{p}{e}\right)^p$, we get: $\Gamma\left(\frac{p}{2}+1\right) \sim \sqrt{2\pi\frac{p}{2}}\left(\frac{p}{2e}\right)^{\frac{p}{2}}$.

1. **Volume of a unit ball in $\mathbb{R}^p$:** $V_p(1) := \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}+1\right)}$.

   where $\Gamma$ is the gamma function, a generalization of the factorial function such that $\Gamma(p) = (p-1)!$ for positive integers $p$.

2. **By Stirling's approximation for large $p$:** That is $p! \approx \sqrt{2\pi p}\left(\frac{p}{e}\right)^p$, we get: $\Gamma\left(\frac{p}{2}+1\right) \sim \sqrt{2\pi\frac{p}{2}}\left(\frac{p}{2e}\right)^{\frac{p}{2}}$.

3. **Asymptotic behavior of $V_p(1)$:** $V_p(1) \sim \left(\frac{2\pi e}{p}\right)^{p/2} (p\pi)^{-1/2}$.

   This shows that the volume of a unit ball in high dimensions becomes very small as $p \to \infty$.

## 2. What can go wrong in high dimensions?
### Volume of a high-dimensional unit ball

1. **Volume of a unit ball in $\mathbb{R}^p$:** $V_p(1) := \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}+1\right)}$.

   where $\Gamma$ is the gamma function, a generalization of the factorial function such that $\Gamma(p) = (p-1)! \quad$ for positive integers $p$.

2. **By Stirling's approximation for large $p$:** That is $p! \approx \sqrt{2\pi p}\left(\frac{p}{e}\right)^p$, we get: $\Gamma\left(\frac{p}{2}+1\right) \sim \sqrt{2\pi\frac{p}{2}}\left(\frac{p}{2e}\right)^{\frac{p}{2}}$.

3. **Asymptotic behavior of $V_p(1)$:** $V_p(1) \sim \left(\frac{2\pi e}{p}\right)^{p/2}(p\pi)^{-1/2}$.

   This shows that the volume of a unit ball in high dimensions becomes very small as $p \to \infty$.

4. **Volume of a ball with radius $r$ in $\mathbb{R}^p$:** $V_p(r) = r^p V_p(1)$.

   This volume also vanishes as $p \to \infty$.

## 2. What can go wrong in high dimensions?
### Unreliable empirical covariance matrix

- In high-dimensional statistics, the empirical covariance matrix can be unreliable.

# 2. What can go wrong in high dimensions?
## Unreliable empirical covariance matrix

- In high-dimensional statistics, the empirical covariance matrix can be unreliable.

### Challenges in High Dimensions

- When $p$ (number of features) is large relative to $n$ (number of observations), the empirical (observed) covariance $\hat{\Sigma}$ may not accurately represent the true (theoretical) covariance.

- In high-dimensional statistics, the empirical covariance matrix can be unreliable.

### Challenges in High Dimensions

- When $p$ (number of features) is large relative to $n$ (number of observations), the empirical (observed) covariance $\hat{\Sigma}$ may not accurately represent the true (theoretical) covariance.
- This is because the estimate can exhibit high variability and may not be invertible.

# 2. What can go wrong in high dimensions?
## Unreliable empirical covariance matrix

- In high-dimensional statistics, the empirical covariance matrix can be unreliable.

### Challenges in High Dimensions

- When $p$ (number of features) is large relative to $n$ (number of observations), the empirical (observed) covariance $\hat{\Sigma}$ may not accurately represent the true (theoretical) covariance.
- This is because the estimate can exhibit high variability and may not be invertible.

- Intuitively, with too many variables and too few samples, the covariance estimate captures random fluctuations more than meaningful patterns.

## 2. What can go wrong in high dimensions?
### Unreliable empirical covariance matrix

- In high-dimensional statistics, the empirical covariance matrix can be unreliable.

**Challenges in High Dimensions**

- When $p$ (number of features) is large relative to $n$ (number of observations), the empirical (observed) covariance $\hat{\Sigma}$ may not accurately represent the true (theoretical) covariance.
- This is because the estimate can exhibit high variability and may not be invertible.

- Intuitively, with too many variables and too few samples, the covariance estimate captures random fluctuations more than meaningful patterns.

**Takeaway**

Be cautious with empirical covariance matrices in high dimensions; alternative techniques or regularization may be needed.

## 2. What can go wrong in high dimensions?
### Unreliable empirical covariance matrix

Empirical covariance in High-Dimension

. Let $x_1, \dots, x_m \overset{iid}{\sim} \mathcal{N}(0, \underbrace{\Sigma}_{p \times p})$ with $\Sigma = I_p$.

. $Sp(\Sigma) = (1, \dots, 1)$ (p times)

. Empirical covariance

$$\hat{\Sigma} = \frac{1}{m} \sum_{i=1}^{m} x_i x_i^T .$$

We have $\operatorname{rank}(\hat{\Sigma}) = m$ so

$$\begin{aligned} m\, \mathbb{E}[|\hat{\Sigma}|_{op}] &\geq \mathbb{E}[Tr(\hat{\Sigma})] \\ &= Tr(\mathbb{E}[\hat{\Sigma}]) \\ &= Tr\Big(\frac{1}{m} \sum_{i=1}^{m} \underbrace{\mathbb{E}[x_i x_i^T]}_{= \Sigma}\Big) \\ &= Tr(\Sigma) = p \end{aligned}$$

So $\mathbb{E}[|\hat{\Sigma}|_{op}] \geq \frac{p}{m} \underset{\text{if } p \gg m}{\gg} 1 = |\Sigma|_{op}$

. Furthermore, we can prove (later) that

$$\mathbb{E}[|\hat{\Sigma}|_{op}] \leq \Big(1 + \sqrt{\frac{p}{m}}\Big)^2 \underset{p \gg m}{=} \frac{p}{m}(1 + o(1))$$

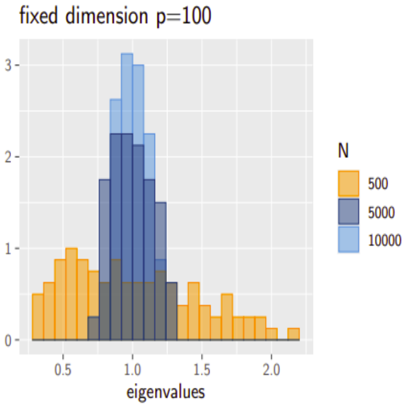
So

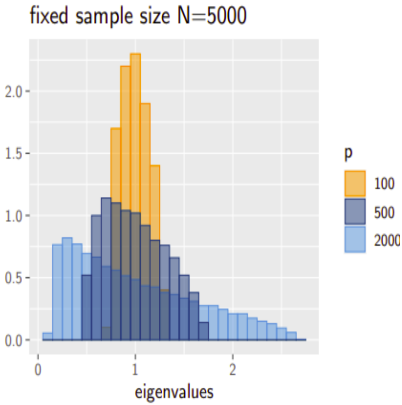
$$Sp(\hat{\Sigma}) \underset{p \gg m}{\approx} \Big(\underbrace{\frac{p}{m}(1+o(1)), \dots, \frac{p}{m}(1+o(1))}_{m \text{ times}}\Big)$$

$\leadsto$ very different from $sp(\Sigma)$,

so we cannot rely on $\hat{\Sigma}$ when $p \gg m$.

## 2. What can go wrong in high dimensions?
### Unreliable empirical covariance matrix




Left: For fixed and small $p$, eigenvalues of $\Sigma$ peak at 1 as $N \to \infty$.
Right: For fixed $N$, eigenvalues of $\hat{\Sigma}$ does not peak at 1 as $p \to \infty$.

## 2. What can go wrong in high dimensions?
### Unbounded distribution of the pairwise distances between points (Lost in high-dimensional spaces)

Let the random variables $X^{(1)}, X^{(2)}, \dots, X^{(n)}$ be i.i.d with $\sim \mathcal{U}([0,1]^p)$ distribution, i.e., independent and uniformly distributed in the hypercube $[0,1]^p$.

**Pairwise Distances:** The distance between two points $X^{(i)}$ and $X^{(j)}$ is given by:

$$d_{ij} = \|X^{(i)} - X^{(j)}\| = \sqrt{\sum_{k=1}^{p} \left(X_k^{(i)} - X_k^{(j)}\right)^2}.$$

Let the random variables $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ be i.i.d with $\sim \mathcal{U}([0,1]^p)$ distribution, i.e., independent and uniformly distributed in the hypercube $[0,1]^p$.

**Pairwise Distances:** The distance between two points $X^{(i)}$ and $X^{(j)}$ is given by:

$$d_{ij} = \|X^{(i)} - X^{(j)}\| = \sqrt{\sum_{k=1}^{p} \left(X_k^{(i)} - X_k^{(j)}\right)^2}.$$

**Expected Value of Squared Distances:**

$$\mathbb{E}\left[\|X^{(i)} - X^{(j)}\|^2\right] = \sum_{k=1}^{p} \mathbb{E}\left[\left(X_k^{(i)} - X_k^{(j)}\right)^2\right] = p \cdot \mathbb{E}\left[(U - U')^2\right] \stackrel{?}{=} \frac{p}{6}.$$

## 2. What can go wrong in high dimensions?
## Unbounded distribution of the pairwise distances between points (Lost in high-dimensional spaces)

Let the random variables $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ be i.i.d with $\sim \mathcal{U}([0,1]^p)$ distribution, i.e., independent and uniformly distributed in the hypercube $[0,1]^p$.

**Pairwise Distances:** The distance between two points $X^{(i)}$ and $X^{(j)}$ is given by:

$$d_{ij} = \|X^{(i)} - X^{(j)}\| = \sqrt{\sum_{k=1}^{p} \left(X_k^{(i)} - X_k^{(j)}\right)^2}.$$

**Expected Value of Squared Distances:**

$$\mathbb{E}\left[\|X^{(i)} - X^{(j)}\|^2\right] = \sum_{k=1}^{p} \mathbb{E}\left[\left(X_k^{(i)} - X_k^{(j)}\right)^2\right] = p \cdot \mathbb{E}\left[(U - U')^2\right] \overset{?}{=} \frac{p}{6}.$$

**Standard Deviation of Squared Distances:**

$$\text{Std}\left[\|X^{(i)} - X^{(j)}\|^2\right] = \sqrt{\sum_{k=1}^{p} \text{Var}\left[\left(X_k^{(i)} - X_k^{(j)}\right)^2\right]} = \sqrt{p \cdot \text{Var}\left[(U - U')^2\right]} \overset{?}{\approx} 0.2\sqrt{p}.$$

where $U$ and $U'$ are two i.i.d. random variables with $\sim \mathcal{U}([0,1])$ distribution.

## 2. What can go wrong in high dimensions?
## Unbounded distribution of the pairwise distances between points
## (Lost in high-dimensional spaces)

Let the random variables $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ be i.i.d with $\sim \mathcal{U}([0,1]^p)$ distribution, i.e., independent and uniformly distributed in the hypercube $[0,1]^p$.

**Pairwise Distances:** The distance between two points $X^{(i)}$ and $X^{(j)}$ is given by:

$$d_{ij} = \|X^{(i)} - X^{(j)}\| = \sqrt{\sum_{k=1}^{p} \left(X_k^{(i)} - X_k^{(j)}\right)^2}.$$

**Expected Value of Squared Distances:**

$$\mathbb{E}\left[\|X^{(i)} - X^{(j)}\|^2\right] = \sum_{k=1}^{p} \mathbb{E}\left[\left(X_k^{(i)} - X_k^{(j)}\right)^2\right] = p \cdot \mathbb{E}\left[(U - U')^2\right] \stackrel{?}{=} \frac{p}{6}.$$

**Standard Deviation of Squared Distances:**

$$\text{Std}\left[\|X^{(i)} - X^{(j)}\|^2\right] = \sqrt{\sum_{k=1}^{p} \text{Var}\left[\left(X_k^{(i)} - X_k^{(j)}\right)^2\right]} = \sqrt{p \cdot \text{Var}\left[(U - U')^2\right]} \stackrel{?}{\approx} 0.2\sqrt{p}.$$

where $U$ and $U'$ are two i.i.d. random variables with $\sim \mathcal{U}([0,1])$ distribution.

As the dimension $p \to \infty$, both the mean and variance of squared distances grow, causing distances to become increasingly large on average. This makes tasks like clustering or nearest neighbor classification challenging, as points become relatively far from each other, making it difficult to define meaningful similarities or groups based on distance metrics.

## Take Home Message (so far)

In high-dimensional spaces, be careful not to be mislead by your low-dimensional intuitions !!

## 2. What can go wrong in high dimensions?
### Curse 1 of Dimensionality: Fluctuations cumulate

Let $X^{(1)}, \dots, X^{(n)} \in \mathbb{R}^p$ be i.i.d. random vectors with $\mathrm{cov}(X) = \sigma^2 I_p$. We want to estimate the expected value $\mathbb{E}[X]$ using the sample mean:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X^{(i)}.$$

Let $X^{(1)}, \dots, X^{(n)} \in \mathbb{R}^p$ be i.i.d. random vectors with $\text{cov}(X) = \sigma^2 I_p$. We want to estimate the expected value $\mathbb{E}[X]$ using the sample mean:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X^{(i)}.$$

**Variance of the Sample Mean:**

$$\mathbb{E}\left[\|\overline{X}_n - \mathbb{E}[X]\|^2\right] = \sum_{j=1}^{p} \mathbb{E}\left[\left(\overline{X}_{n,j} - \mathbb{E}[X_j]\right)^2\right] = \sum_{j=1}^{p} \text{Var}(\overline{X}_{n,j}).$$

Since $\text{Var}(\overline{X}_{n,j}) = \frac{\sigma^2}{n}$, we get:

$$\mathbb{E}\left[\|\overline{X}_n - \mathbb{E}[X]\|^2\right] = \frac{p\sigma^2}{n}.$$

Let $X^{(1)}, \dots, X^{(n)} \in \mathbb{R}^p$ be i.i.d. random vectors with $\mathrm{cov}(X) = \sigma^2 I_p$. We want to estimate the expected value $\mathbb{E}[X]$ using the sample mean:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X^{(i)}.$$

**Variance of the Sample Mean:**

$$\mathbb{E}\left[\|\overline{X}_n - \mathbb{E}[X]\|^2\right] = \sum_{j=1}^{p} \mathbb{E}\left[\left(\overline{X}_{n,j} - \mathbb{E}[X_j]\right)^2\right] = \sum_{j=1}^{p} \mathrm{Var}(\overline{X}_{n,j}).$$

Since $\mathrm{Var}(\overline{X}_{n,j}) = \frac{\sigma^2}{n}$, we get:

$$\mathbb{E}\left[\|\overline{X}_n - \mathbb{E}[X]\|^2\right] = \frac{p\sigma^2}{n}.$$

**Implication:** When $p \gg n$, the error in estimating the mean grows with the dimensionality $p$, making it difficult to accurately estimate $\mathbb{E}[X]$.

**Observations:** $(Y_i, X^{(i)}) \in \mathbb{R} \times [0,1]^p$ for $i = 1, \ldots, n$.

**Model:** $Y_i = f(X^{(i)}) + \epsilon_i$, where $f$ is smooth and the $\epsilon_i$ are noise terms.

Assume that $(Y_i, X^{(i)})_{i=1,\ldots,n}$ are i.i.d., and that $X^{(i)} \sim \mathcal{U}([0,1]^p)$.

**Local averaging:** Estimate $f$ using local averaging:

$$\hat{f}(x) = \text{average of } \{Y_i : X^{(i)} \text{ close to } x\}.$$

**Observations:** $(Y_i, X^{(i)}) \in \mathbb{R} \times [0,1]^p$ for $i = 1, \dots, n$.

**Model:** $Y_i = f(X^{(i)}) + \epsilon_i$, where $f$ is smooth and the $\epsilon_i$ are noise terms.

Assume that $(Y_i, X^{(i)})_{i=1,\dots,n}$ are i.i.d., and that $X^{(i)} \sim \mathcal{U}([0,1]^p)$.

**Local averaging:** Estimate $f$ using local averaging:

$$\hat{f}(x) = \text{average of } \{Y_i : X^{(i)} \text{ close to } x\}.$$

- The **union bound** (Boole's Inequality) states that for any events $A_1, A_2, \dots, A_n$:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

## 2. What can go wrong in high dimensions?
### Curse 2 of Dimensionality: Local averaging is ineffective

**Observations:** $(Y_i, X^{(i)}) \in \mathbb{R} \times [0,1]^p$ for $i = 1, \ldots, n$.

**Model:** $Y_i = f(X^{(i)}) + \epsilon_i$, where $f$ is smooth and the $\epsilon_i$ are noise terms.

Assume that $(Y_i, X^{(i)})_{i=1,\ldots,n}$ are i.i.d., and that $X^{(i)} \sim \mathcal{U}([0,1]^p)$.

**Local averaging:** Estimate $f$ using local averaging:

$$\hat{f}(x) = \text{average of } \{Y_i : X^{(i)} \text{ close to } x\}.$$

- The **union bound** (Boole's Inequality) states that for any events $A_1, A_2, \ldots, A_n$:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

- The event $\exists i \in \{1, \ldots, n\} : \|x - X^{(i)}\| \leq \delta$ is equivalent to the union of the events $A_i$:

$$\exists i \in \{1, \ldots, n\} : \|x - X^{(i)}\| \leq \delta \quad \Leftrightarrow \quad \bigcup_{i=1}^{n} \left\{\|x - X^{(i)}\| \leq \delta\right\}.$$

- Applying the union bound to this union of events gives:

$$\mathbb{P}\left[\bigcup_{i=1}^{n}\left\{\|x - X^{(i)}\| \leq \delta\right\}\right] \leq \sum_{i=1}^{n}\mathbb{P}\left[\|x - X^{(i)}\| \leq \delta\right].$$

- Applying the union bound to this union of events gives:

$$\mathbb{P}\left[\bigcup_{i=1}^{n}\left\{\|x - X^{(i)}\| \leq \delta\right\}\right] \leq \sum_{i=1}^{n} \mathbb{P}\left[\|x - X^{(i)}\| \leq \delta\right].$$

- Since the $X^{(i)} \overset{i.i.d}{\sim} \mathcal{U}([0,1]^p)$ , each term in the sum is equal:

$$\mathbb{P}\left[\|x - X^{(1)}\| \leq \delta\right] = \mathbb{P}\left[\|x - X^{(2)}\| \leq \delta\right] = \ldots = \mathbb{P}\left[\|x - X^{(n)}\| \leq \delta\right].$$

## 2. What can go wrong in high dimensions?
### Curse 2 of Dimensionality: Local averaging is ineffective

- Applying the union bound to this union of events gives:

$$\mathbb{P}\left[\bigcup_{i=1}^{n}\left\{\|x - X^{(i)}\| \leq \delta\right\}\right] \leq \sum_{i=1}^{n} \mathbb{P}\left[\|x - X^{(i)}\| \leq \delta\right].$$

- Since the $X^{(i)} \overset{i.i.d}{\sim} \mathcal{U}([0,1]^p)$ , each term in the sum is equal:

$$\mathbb{P}\left[\|x - X^{(1)}\| \leq \delta\right] = \mathbb{P}\left[\|x - X^{(2)}\| \leq \delta\right] = \ldots = \mathbb{P}\left[\|x - X^{(n)}\| \leq \delta\right].$$

- Thus, the sum simplifies to:

$$\sum_{i=1}^{n} \mathbb{P}\left[\|x - X^{(i)}\| \leq \delta\right] = n\,\mathbb{P}\left[\|x - X^{(1)}\| \leq \delta\right].$$

**Problem:** For $x \in [0,1]^p$, we have:

$$\mathbb{P}\left[\exists i \in \{1, \ldots, n\} : \|x - X^{(i)}\| \leq \delta\right] \leq n\mathbb{P}\left[\|x - X^{(1)}\| \leq \delta\right] \leq nV_p(\delta),$$

where $V_p(\delta)$ is the volume of a ball of radius $\delta$ in $\mathbb{R}^p$.

## 2. What can go wrong in high dimensions?
### Curse 2 of Dimensionality: Local averaging is ineffective

As $p \to \infty$: The volume $V_p(\delta) \approx \left(\frac{2\pi e}{p}\right)^{p/2} \delta^p \sqrt{\pi p}$, and the probability decreases rapidly:

$$nV_p(\delta) \to 0 \text{ as } p \to \infty.$$

As $p \to \infty$: The volume $V_p(\delta) \approx \left(\frac{2\pi e}{p}\right)^{p/2} \delta^p \sqrt{\pi p}$, and the probability decreases rapidly:

$$nV_p(\delta) \to 0 \text{ as } p \to \infty.$$

**Conclusion:** In high dimensions, finding enough data points close to $x$ becomes extremely unlikely, making local averaging ineffective.

As $p \to \infty$: The volume $V_p(\delta) \approx \left(\frac{2\pi e}{p}\right)^{p/2} \delta^p \sqrt{\pi p}$, and the probability decreases rapidly:

$$nV_p(\delta) \to 0 \text{ as } p \to \infty.$$

**Conclusion:** In high dimensions, finding enough data points close to $x$ becomes extremely unlikely, making local averaging ineffective.

**Question: Which sample size is needed to avoid the loss of locality?**

Number $n$ of points $x_1, \ldots, x_n$ required for having at least one observation at distance $\delta = 1$ with probability $1/2$:

$$n \geq \frac{1}{2V_p(1)} \quad \text{where} \quad V_p(1) = \text{Volume of a unit ball in } \mathbb{R}^p.$$

As $p \to \infty$: The volume $V_p(\delta) \approx \left(\frac{2\pi e}{p}\right)^{p/2} \delta^p \sqrt{\pi p}$, and the probability decreases rapidly:

$$nV_p(\delta) \to 0 \text{ as } p \to \infty.$$

**Conclusion:** In high dimensions, finding enough data points close to $x$ becomes extremely unlikely, making local averaging ineffective.

**Question: Which sample size is needed to avoid the loss of locality?**

Number $n$ of points $x_1, \ldots, x_n$ required for having at least one observation at distance $\delta = 1$ with probability $1/2$:

$$n \geq \frac{1}{2V_p(1)} \quad \text{where} \quad V_p(1) = \text{Volume of a unit ball in } \mathbb{R}^p.$$

**Asymptotic Behavior:** For large $p$, we have: $V_p(1) \sim \left(\frac{2\pi e}{p}\right)^{p/2} \sqrt{\pi p}$.
**Implications for Sample Size:**

- As $p \to \infty$, $V_p(1)$ decreases rapidly, leading to: $n \geq \left(\frac{p}{2\pi e}\right)^{p/2} \sqrt{\frac{p\pi}{4}}$.

### Which sample size is needed to avoid the loss of locality?

**Example:** For different values of $p$:

| Dimension $p$ | Estimated Sample Size $n$ |
|---|---|
| 20 | 39 |
| 30 | 45630 |
| 50 | $5.7 \times 10^{12}$ |
| 100 | $4.2 \times 10^{39}$ |
| 200 | Larger than the estimated number of particles in the observable universe |

- This required sample size grows exponentially with $p$, making it impractically large for high dimensions.

**Threshold for Detection Increases:**

- In high-dimensional spaces, detecting a weak signal $\theta_j$ becomes challenging because the signal must be significantly larger than the noise.
- Suppose $Z_j \sim N(\sqrt{n}\theta_j, \sigma^2)$, where:
  - $n$: Number of observations.
  - $\sigma^2$: Variance of the noise.

**Threshold for Detection Increases:**

- In high-dimensional spaces, detecting a weak signal $\theta_j$ becomes challenging because the signal must be significantly larger than the noise.
- Suppose $Z_j \sim N(\sqrt{n}\theta_j, \sigma^2)$, where:
  - $n$: Number of observations.
  - $\sigma^2$: Variance of the noise.
- In 1D, the threshold for detecting a signal is $|\theta_j| \geq \frac{2\sigma}{\sqrt{n}}$

**Threshold for Detection Increases:**

- In high-dimensional spaces, detecting a weak signal $\theta_j$ becomes challenging because the signal must be significantly larger than the noise.
- Suppose $Z_j \sim N(\sqrt{n}\theta_j, \sigma^2)$, where:
  - $n$: Number of observations.
  - $\sigma^2$: Variance of the noise.
- In 1D, the threshold for detecting a signal is $|\theta_j| \geq \frac{2\sigma}{\sqrt{n}}$
- For $W_1, ..., W_p \overset{i.i.d}{\sim} \mathcal{N}([0,1]^p)$, we have $\max\limits_{j=1,..,p} W_j^2 = 2\sigma^2 \log(p)$,

**Threshold for Detection Increases:**

- In high-dimensional spaces, detecting a weak signal $\theta_j$ becomes challenging because the signal must be significantly larger than the noise.
- Suppose $Z_j \sim N(\sqrt{n}\theta_j, \sigma^2)$, where:
  - $n$: Number of observations.
  - $\sigma^2$: Variance of the noise.
- In 1D, the threshold for detecting a signal is $|\theta_j| \geq \frac{2\sigma}{\sqrt{n}}$
- For $W_1, ..., W_p \overset{i.i.d}{\sim} \mathcal{N}([0,1]^p)$, we have $\max\limits_{j=1,..,p} W_j^2 = 2\sigma^2 \log(p)$,
- In high dimensions $p$, the threshold for detecting a signal increases with the dimensionality:

$$|\theta_j| \geq \sigma\sqrt{\frac{2\log(p)}{n}}.$$

**Threshold for Detection Increases:**

- In high-dimensional spaces, detecting a weak signal $\theta_j$ becomes challenging because the signal must be significantly larger than the noise.
- Suppose $Z_j \sim N(\sqrt{n}\theta_j, \sigma^2)$, where:
  - $n$: Number of observations.
  - $\sigma^2$: Variance of the noise.
- In 1D, the threshold for detecting a signal is $|\theta_j| \geq \frac{2\sigma}{\sqrt{n}}$
- For $W_1, ..., W_p \overset{i.i.d}{\sim} \mathcal{N}([0,1]^p)$, we have $\max_{j=1,..,p} W_j^2 = 2\sigma^2 \log(p)$,
- In high dimensions $p$, the threshold for detecting a signal increases with the dimensionality:

$$|\theta_j| \geq \sigma\sqrt{\frac{2\log(p)}{n}}.$$

- As $p$ increases, the signal $\theta_j$ needs to be stronger (larger) to stand out against the noise.
- This makes it harder for weak signals to be detected, as they become obscured by the noise in high-dimensional data.

## Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

In high-dimensional data, where the number of predictors $p$ is large, linear regression faces significant computational challenges. These challenges can affect both the feasibility and stability of the model.

The linear regression model is defined as:

$$y = X\beta + \epsilon,$$

where: $y \in \mathbb{R}^n$ (response vector), $X \in \mathbb{R}^{n\times p}$ (design matrix with $n$ observations and $p$ predictors), $\beta \in \mathbb{R}^p$ (coefficient vector) and $\epsilon \in \mathbb{R}^n$ (error vector).

# 2. What can go wrong in high dimensions?
## Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

In high-dimensional data, where the number of predictors $p$ is large, linear regression faces significant computational challenges. These challenges can affect both the feasibility and stability of the model.

The linear regression model is defined as:

$$y = X\beta + \epsilon,$$

where: $y \in \mathbb{R}^n$ (response vector), $X \in \mathbb{R}^{n \times p}$ (design matrix with $n$ observations and $p$ predictors), $\beta \in \mathbb{R}^p$ (coefficient vector) and $\epsilon \in \mathbb{R}^n$ (error vector).

The ordinary least squares (OLS) solution for estimating $\beta$ minimizes the residual sum of squares:

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2.$$

The solution for linear regression coefficients is given by:

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

## Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

In high-dimensional data, where the number of predictors $p$ is large, linear regression faces significant computational challenges. These challenges can affect both the feasibility and stability of the model.

The linear regression model is defined as:

$$y = X\beta + \epsilon,$$

where: $y \in \mathbb{R}^n$ (response vector), $X \in \mathbb{R}^{n \times p}$ (design matrix with $n$ observations and $p$ predictors), $\beta \in \mathbb{R}^p$ (coefficient vector) and $\epsilon \in \mathbb{R}^n$ (error vector).

The ordinary least squares (OLS) solution for estimating $\beta$ minimizes the residual sum of squares:

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2.$$

The solution for linear regression coefficients is given by:

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

**1. Size and inversion of the Matrix $(X^T X)$:**

- The matrix $X^T X$ is a $p \times p$ matrix, where $p$ is the number of predictors.
- As $p$ increases, the size of $X^T X$ increases quadratically ($p^2$ elements).

## Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

In high-dimensional data, where the number of predictors $p$ is large, linear regression faces significant computational challenges. These challenges can affect both the feasibility and stability of the model.

The linear regression model is defined as:

$$y = X\beta + \epsilon,$$

where: $y \in \mathbb{R}^n$ (response vector), $X \in \mathbb{R}^{n \times p}$ (design matrix with $n$ observations and $p$ predictors), $\beta \in \mathbb{R}^p$ (coefficient vector) and $\epsilon \in \mathbb{R}^n$ (error vector).

The ordinary least squares (OLS) solution for estimating $\beta$ minimizes the residual sum of squares:

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2.$$

The solution for linear regression coefficients is given by:

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

**1. Size and inversion of the Matrix $(X^T X)$:**

- The matrix $X^T X$ is a $p \times p$ matrix, where $p$ is the number of predictors.
- As $p$ increases, the size of $X^T X$ increases quadratically ($p^2$ elements).
- As $p$ (number of predictors) increases, the matrix $X^T X$ becomes larger ($p \times p$).
- Computing the inverse of a $p \times p$ matrix has a time complexity of $O(p^3)$, which can become prohibitive when $p$ is large.

# Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

In high-dimensional data, where the number of predictors $p$ is large, linear regression faces significant computational challenges. These challenges can affect both the feasibility and stability of the model.

The linear regression model is defined as:

$$y = X\beta + \epsilon,$$

where: $y \in \mathbb{R}^n$ (response vector), $X \in \mathbb{R}^{n \times p}$ (design matrix with $n$ observations and $p$ predictors), $\beta \in \mathbb{R}^p$ (coefficient vector) and $\epsilon \in \mathbb{R}^n$ (error vector).

The ordinary least squares (OLS) solution for estimating $\beta$ minimizes the residual sum of squares:

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2.$$

The solution for linear regression coefficients is given by:

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

**1. Size and inversion of the Matrix $(X^T X)$:**

- The matrix $X^T X$ is a $p \times p$ matrix, where $p$ is the number of predictors.
- As $p$ increases, the size of $X^T X$ increases quadratically ($p^2$ elements).
- As $p$ (number of predictors) increases, the matrix $X^T X$ becomes larger ($p \times p$).
- Computing the inverse of a $p \times p$ matrix has a time complexity of $O(p^3)$, which can become prohibitive when $p$ is large.
- When $p$ is large, the $O(p^3)$ complexity becomes computationally expensive, making the inversion step a bottleneck.
- This complexity arises from performing *Gaussian elimination* or using other matrix decomposition methods (e.g., *LU decomposition*).

**2. Numerical Stability:**

- When $p$ is large, multicollinearity (high correlation between predictors) is common.

## 2. What can go wrong in high dimensions?
### Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

**2. Numerical Stability:**

- When $p$ is large, multicollinearity (high correlation between predictors) is common.
- When predictors are highly correlated, the columns of the design matrix $X$ become nearly linearly dependent.

## Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

**2. Numerical Stability:**

- When $p$ is large, multicollinearity (high correlation between predictors) is common.
- When predictors are highly correlated, the columns of the design matrix $X$ become nearly linearly dependent.
- Multicollinearity makes $X^T X$ nearly singular (non-invertible), causing issues with numerical stability.

## 2. What can go wrong in high dimensions?
### Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

**2. Numerical Stability:**

- When $p$ is large, multicollinearity (high correlation between predictors) is common.
- When predictors are highly correlated, the columns of the design matrix $X$ become nearly linearly dependent.
- Multicollinearity makes $X^T X$ nearly singular (non-invertible), causing issues with numerical stability.
- Inverting a nearly singular matrix can result in large computational errors, leading to unstable coefficient estimates.

## Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

**2. Numerical Stability:**

- When $p$ is large, multicollinearity (high correlation between predictors) is common.
- When predictors are highly correlated, the columns of the design matrix $X$ become nearly linearly dependent.
- Multicollinearity makes $X^T X$ nearly singular (non-invertible), causing issues with numerical stability.
- Inverting a nearly singular matrix can result in large computational errors, leading to unstable coefficient estimates.
- Even small changes in $X$ or $y$ can cause large variations in $\hat{\beta}$, leading to unreliable estimates of the regression coefficients, making it difficult to draw accurate conclusions from the linear regression model.

**2. Numerical Stability:**

- When $p$ is large, multicollinearity (high correlation between predictors) is common.
- When predictors are highly correlated, the columns of the design matrix $X$ become nearly linearly dependent.
- Multicollinearity makes $X^T X$ nearly singular (non-invertible), causing issues with numerical stability.
- Inverting a nearly singular matrix can result in large computational errors, leading to unstable coefficient estimates.
- Even small changes in $X$ or $y$ can cause large variations in $\hat{\beta}$, leading to unreliable estimates of the regression coefficients, making it difficult to draw accurate conclusions from the linear regression model.
- Regularization methods (e.g., Ridge regression) add a penalty term to $X^T X$, making it better conditioned and more stable to invert.

**4. Computational Complexity of Algorithms:**

- As $p$ increases, algorithms that solve for $\hat{\beta}$ become slower due to the increased complexity of operations like matrix multiplication and inversion.

# 2. What can go wrong in high dimensions?
## Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

**2. Numerical Stability:**

- When $p$ is large, multicollinearity (high correlation between predictors) is common.
- When predictors are highly correlated, the columns of the design matrix $X$ become nearly linearly dependent.
- Multicollinearity makes $X^T X$ nearly singular (non-invertible), causing issues with numerical stability.
- Inverting a nearly singular matrix can result in large computational errors, leading to unstable coefficient estimates.
- Even small changes in $X$ or $y$ can cause large variations in $\hat{\beta}$, leading to unreliable estimates of the regression coefficients, making it difficult to draw accurate conclusions from the linear regression model.
- Regularization methods (e.g., Ridge regression) add a penalty term to $X^T X$, making it better conditioned and more stable to invert.

**4. Computational Complexity of Algorithms:**

- As $p$ increases, algorithms that solve for $\hat{\beta}$ become slower due to the increased complexity of operations like matrix multiplication and inversion.
- Iterative methods, such as gradient descent, can be used as an alternative to direct inversion but may require many iterations to converge in high dimensions.

# 2. What can go wrong in high dimensions?
## Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

**2. Numerical Stability:**
- When $p$ is large, multicollinearity (high correlation between predictors) is common.
- When predictors are highly correlated, the columns of the design matrix $X$ become nearly linearly dependent.
- Multicollinearity makes $X^T X$ nearly singular (non-invertible), causing issues with numerical stability.
- Inverting a nearly singular matrix can result in large computational errors, leading to unstable coefficient estimates.
- Even small changes in $X$ or $y$ can cause large variations in $\hat{\beta}$, leading to unreliable estimates of the regression coefficients, making it difficult to draw accurate conclusions from the linear regression model.
- Regularization methods (e.g., Ridge regression) add a penalty term to $X^T X$, making it better conditioned and more stable to invert.

**4. Computational Complexity of Algorithms:**
- As $p$ increases, algorithms that solve for $\hat{\beta}$ become slower due to the increased complexity of operations like matrix multiplication and inversion.
- Iterative methods, such as gradient descent, can be used as an alternative to direct inversion but may require many iterations to converge in high dimensions.
- Regularization techniques like Lasso and Ridge reduce the complexity but require optimization procedures that can be computationally intensive.

**5. Practical Implications:**
- High-dimensional linear regression may become infeasible without sufficient computational resources.

# 2. What can go wrong in high dimensions?
## Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

**2. Numerical Stability:**

- When $p$ is large, multicollinearity (high correlation between predictors) is common.
- When predictors are highly correlated, the columns of the design matrix $X$ become nearly linearly dependent.
- Multicollinearity makes $X^T X$ nearly singular (non-invertible), causing issues with numerical stability.
- Inverting a nearly singular matrix can result in large computational errors, leading to unstable coefficient estimates.
- Even small changes in $X$ or $y$ can cause large variations in $\hat{\beta}$, leading to unreliable estimates of the regression coefficients, making it difficult to draw accurate conclusions from the linear regression model.
- Regularization methods (e.g., Ridge regression) add a penalty term to $X^T X$, making it better conditioned and more stable to invert.

**4. Computational Complexity of Algorithms:**

- As $p$ increases, algorithms that solve for $\hat{\beta}$ become slower due to the increased complexity of operations like matrix multiplication and inversion.
- Iterative methods, such as gradient descent, can be used as an alternative to direct inversion but may require many iterations to converge in high dimensions.
- Regularization techniques like Lasso and Ridge reduce the complexity but require optimization procedures that can be computationally intensive.

**5. Practical Implications:**

- High-dimensional linear regression may become infeasible without sufficient computational resources.
- Optimized algorithms and specialized hardware (e.g., GPUs) can help mitigate some of these challenges.

## 2. What can go wrong in high dimensions?
### Curse 4 of Dimensionality: Multicollinearity (ill-conditioning) of matrices

**2. Numerical Stability:**

- When $p$ is large, multicollinearity (high correlation between predictors) is common.
- When predictors are highly correlated, the columns of the design matrix $X$ become nearly linearly dependent.
- Multicollinearity makes $X^T X$ nearly singular (non-invertible), causing issues with numerical stability.
- Inverting a nearly singular matrix can result in large computational errors, leading to unstable coefficient estimates.
- Even small changes in $X$ or $y$ can cause large variations in $\hat{\beta}$, leading to unreliable estimates of the regression coefficients, making it difficult to draw accurate conclusions from the linear regression model.
- Regularization methods (e.g., Ridge regression) add a penalty term to $X^T X$, making it better conditioned and more stable to invert.

**4. Computational Complexity of Algorithms:**

- As $p$ increases, algorithms that solve for $\hat{\beta}$ become slower due to the increased complexity of operations like matrix multiplication and inversion.
- Iterative methods, such as gradient descent, can be used as an alternative to direct inversion but may require many iterations to converge in high dimensions.
- Regularization techniques like Lasso and Ridge reduce the complexity but require optimization procedures that can be computationally intensive.

**5. Practical Implications:**

- High-dimensional linear regression may become infeasible without sufficient computational resources.
- Optimized algorithms and specialized hardware (e.g., GPUs) can help mitigate some of these challenges.
- Reducing the dimensionality of data through methods like PCA before fitting the model can make computations more manageable.

## 2. What can go wrong in high dimensions?
### Some other curses of dimensionality

- Curse 5: An accumulation of rare events may not be rare (false discoveries, etc). In high-dimensional data, each dimension or variable can be associated with a certain type of "rare event." For example, detecting an anomaly in one variable might be rare, but when you have many variables, the chances that at least one of these variables will exhibit a rare event increase. This is because the probability of encountering at least one rare event grows with the number of variables.

## 2. What can go wrong in high dimensions?
### Some other curses of dimensionality

- Curse 5: An accumulation of rare events may not be rare (false discoveries, etc). In high-dimensional data, each dimension or variable can be associated with a certain type of "rare event." For example, detecting an anomaly in one variable might be rare, but when you have many variables, the chances that at least one of these variables will exhibit a rare event increase. This is because the probability of encountering at least one rare event grows with the number of variables.
- Curse 6: Algorithmic complexity must remain low. When $p$ is large, an algorithmic complexity larger than $O(p^2)$ is computationally prohibitive. For very large $p$, even a complexity $O(p^2)$ can be an issue.
- etc

**Hopeless?**

High-dimensional data are often concentrated around low-dimensional structures, reflecting the (relatively) small complexity of the systems producing the data. Examples of low-dimensional structures:

- Geometrical structures in images.
- Regulation networks of a biological system.
- Social structures in marketing data.
- Human technologies have limited complexity, etc

**Dimension Reduction:**

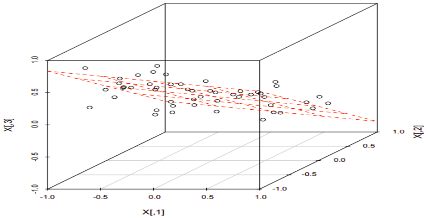
- **Unsupervised**: Principal Component Analysis (PCA).
- **Supervised**: Methods that utilize labeled data.

For any data points $X^{(1)}, \ldots, X^{(n)} \in \mathbb{R}^p$ and any dimension $d \leq p$, PCA computes the linear subspace $V_d$ such that:

$$V_d \in \arg \min_{\mathrm{dim}(V) \leq d} \sum_{i=1}^{n} \|X^{(i)} - \mathrm{Proj}_V X^{(i)}\|^2,$$

where $\mathrm{Proj}_V$ is the orthogonal projection matrix onto the subspace $V$.

**Example:** $V_2$ in dimension $p = 3$.




**Recap on PCA:** PCA finds a lower-dimensional representation of the data that preserves the directions with the most variance, allowing us to approximate the data using fewer dimensions.

# 4. Summary

**What are the different viewpoints?**

- Classical asymptotics.
- High-dimensional asymptotics.
- Non-asymptotic bounds.

# 4. Summary

**What are the different viewpoints?**

- Classical asymptotics.
- High-dimensional asymptotics.
- Non-asymptotic bounds.

**What can go wrong in high dimensions?**

- no consistent estimator
- low-rank matrices, not invertible

# 4. Summary

**What are the different viewpoints?**

- Classical asymptotics.
- High-dimensional asymptotics.
- Non-asymptotic bounds.

**What can go wrong in high dimensions?**

- no consistent estimator
- low-rank matrices, not invertible

**What can help?**

- Finding or imposing lower dimensional structure
- sparsity
- low rank
- graphical structure

# 4. Summary

**What are the different viewpoints?**

- Classical asymptotics.
- High-dimensional asymptotics.
- Non-asymptotic bounds.

**What can go wrong in high dimensions?**

- no consistent estimator
- low-rank matrices, not invertible

**What can help?**

- Finding or imposing lower dimensional structure
- sparsity
- low rank
- graphical structure

Key questions:
What embedded low-dimensional structures are present in data?
How can they be exploited?