Statistics: The Backbone of Statistical Learning

Alexander G. Ororbia II and Viet Nguyen

Introduction to Machine Learning

CSCI-335

2/6/2026

Statistics

Standard deviation - a measure of data points differ from mean

Average of differences (w/ mean as reference point)

Higher standard deviation indicates higher spread, less consistency, and less "clustering/blobbing"

Sample standard deviation: $s = \sqrt{\frac{\sum\left(x - \bar{X}\right)^2}{n-1}}$

Mean (average): $\bar{X} = \left(\frac{1}{n}\right)\sum(x)$

Population standard deviation: $\sigma = \sqrt{\frac{\sum\left(x - \mu\right)^2}{N}}$

Expectation

X is univariate here!

$$\mathrm{E}[X] = \sum_{i=1}^{k} x_i\, p_i = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$$

Definition A *random vector* $\vec{X}$ is a vector $(X_1, X_2, \ldots, X_p)$ of jointly distributed random variables. As is customary in linear algebra, we will write vectors as column matrices whenever convenient.

Definition The expectation $E\vec{X}$ of a random vector $\vec{X} = [X_1, X_2, \ldots, X_p]^T$ is given by

$$E\vec{X} = \begin{bmatrix} EX_1 \\ EX_2 \\ \vdots \\ EX_p \end{bmatrix}.$$

$$\begin{aligned} E\vec{X} &= E\begin{bmatrix} X_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + E\begin{bmatrix} 0 \\ X_2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + E\begin{bmatrix} 0 \\ 0 \\ \vdots \\ X_p \end{bmatrix} \\ &= \begin{bmatrix} EX_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ EX_2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ EX_p \end{bmatrix} \\ &= \begin{bmatrix} EX_1 \\ EX_2 \\ \vdots \\ EX_p \end{bmatrix}. \end{aligned}$$

The linearity properties of the expectation can be expressed compactly by stating that for any $k \times p$-matrix A and any $1 \times j$-matrix B,

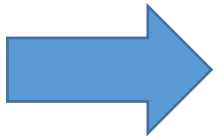
$$E(A\vec{X}) = AE\vec{X} \quad \text{and} \quad E(\vec{X}B) = (E\vec{X})B.$$

X is *p x 1* here!

X is *j x 1* here!

Covariance

- Variance and Covariance:
  - Measure of the “spread” of a set of points around their center of mass (mean)
- Variance:
  - Measure of deviation from mean for points in one dimension
- Covariance:
  - Measure of how much each of dimensions vary from mean with **respect to each other**



- **Covariance is measured between two dimensions**
- **Covariance → relation between two dimensions**
- **Covariance between one dimension is variance**

Variance-Covariance

Definition The *variance–covariance matrix* (or simply the *covariance matrix*) of a random vector $\vec{X}$ is given by:

$$\mathrm{Cov}(\vec{X}) = E\left[(\vec{X} - E\vec{X})(\vec{X} - E\vec{X})^T\right].$$

Proposition

$$\mathrm{Cov}(\vec{X}) = E[\vec{X}\vec{X}^T] - E\vec{X}(E\vec{X})^T.$$

Proposition

$$\mathrm{Cov}(\vec{X}) = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_p) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \mathrm{Cov}(X_p, X_2) & \cdots & \mathrm{Var}(X_p) \end{bmatrix}.$$

Thus, $\mathrm{Cov}(\vec{X})$ *is a symmetric matrix, since* $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

Covariance Matrix

- Gives a measure of the dispersion of the data
- It is a D **x** D matrix
  - Element in position i,j is the covariance between the i^{th} and j^{th} variables.
    - Covariance between two variables x_i and x_j is defined as $E[(x_i-\mu_i)(y_i-\mu_j)]$
    - Can be positive or negative
  - If the variables are independent then the covariance is zero.
    - Then all matrix elements are zero except diagonal elements which represent the variances

The Gaussian Distribution

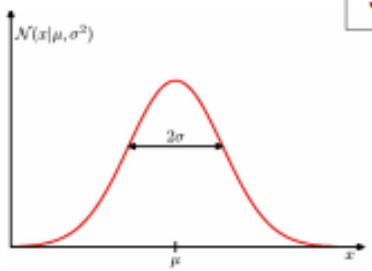
Carl Friedrich Gauss
1777-1855

- For single real-valued variable x

$$N(x \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

68% of data lies within σ of mean
95% within 2σ

- Parameters:
  - Mean μ, variance σ^2,
- *Standard deviation* σ



- *Precision* $\beta = 1/\sigma^2$, $E[x] = \mu$, $Var[x] = \sigma^2$
- For D-dimensional vector $\mathbf{x}$, multivariate Gaussian

$$N(\mathbf{x} \mid \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right\}$$

$\boldsymbol{\mu}$ is a mean vector, Σ is a D x D covariance matrix, $|\Sigma|$ is the determinant of Σ

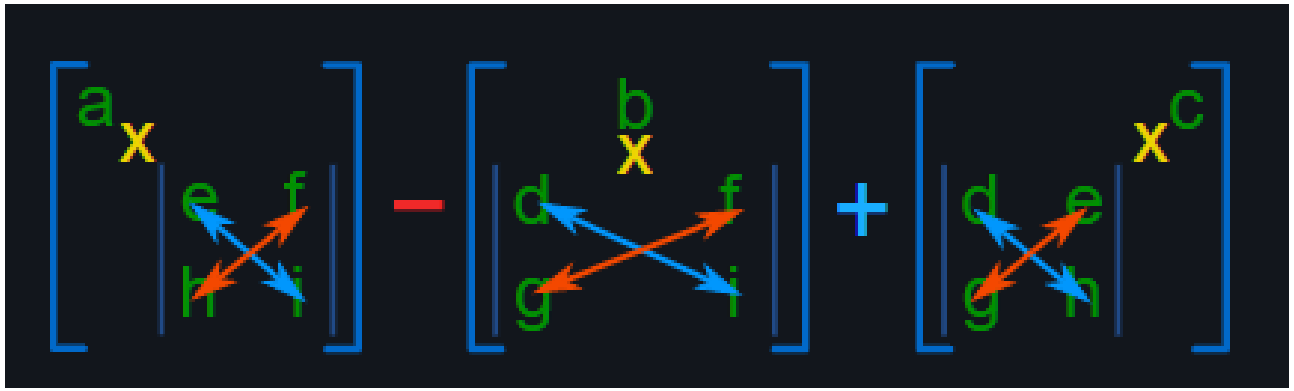
Σ^{-1} is also referred to as the precision matrix

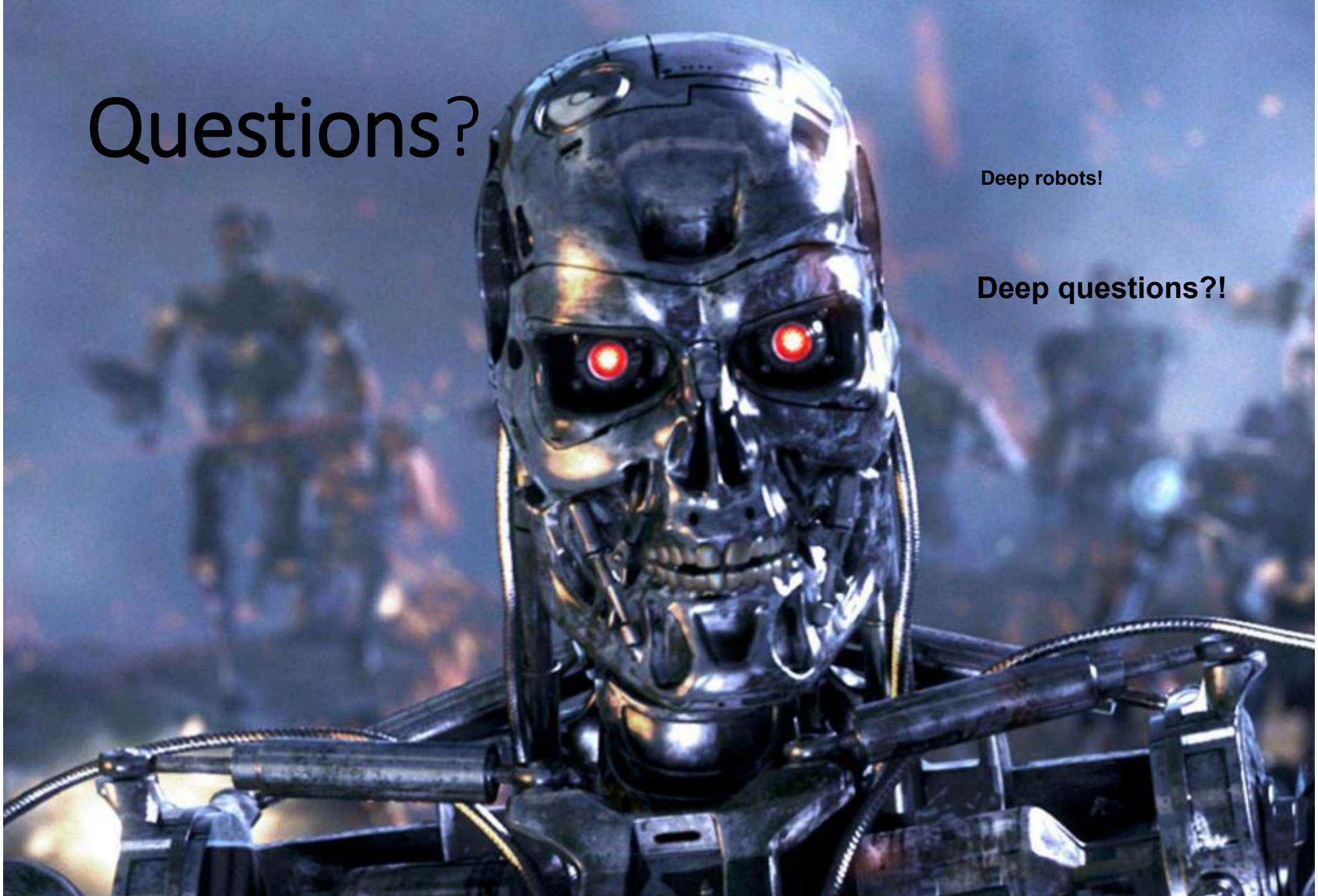
Matrix Determinant

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

$$|A| = a(ei - fh) - b(di - fg) + c(dh - eg)$$

"The determinant of A equals ... etc"




Questions?
Deep robots!
Deep questions?!