Foundations of Machine Learning (CS 725)

FALL 2024

Lecture 2:

- Introduction to Linear Regression

Instructor: Preethi Jyothi

Question 1a

Say you have a continuous random variable X with pdf $g(x)$.

True or False? $\int_{-\infty}^{\infty} g(x)dx = 1$

Question 1b

Say you have a continuous random variable X with pdf $g(x)$.

True or False? $g(x)$ can be greater than 1

Question 1c

Say you have a continuous random variable X with pdf $g(x)$.

True or False? $\int_a^b g(x)dx$ can be greater than 1

Question 2

We are given two random variables X and Y. X is discrete and can take values $-2, -1, 0, 1, 2,$ with probability $\frac{1}{5}$ each. Let $Y = X^2$. What is the covariance of X and Y, $\mathrm{Cov}[X, Y]$?

Ans: 0

Recap: Learning a Predictor Function

Consider a target/true function $f : \mathcal{X} \rightarrow \mathcal{Y}$ that holds over a training dataset $\mathcal{D} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\}, \mathbf{x}_i \in \mathcal{X}, y_i \in \mathcal{Y}$. Our goal is to find a predictor function or hypothesis $h : \mathcal{X} \rightarrow \mathcal{Y}$ that closely approximates f.

1. What functions are permissible for the hypothesis h? [**Hypothesis Class**]

2. How can we quantify the performance of the hypothesis? [**Loss/Error Function**]

3. How do we find the best hypothesis? [**Optimization**]

Linear Regression

1. What functions are permissible for the hypothesis h? [**Hypothesis Class**]

   Hypothesis class $\mathcal{H}$ is: $\{h_{\mathbf{w}} : h_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^T\mathbf{x}, \mathbf{w} \in \mathbb{R}^{d+1}\}$

2. How can we quantify the performance of the hypothesis? [**Loss/Error Function**]

   Least squares (or mean squared) loss: $\mathcal{L}_{\text{MSE}} = \sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x}_i)^2$

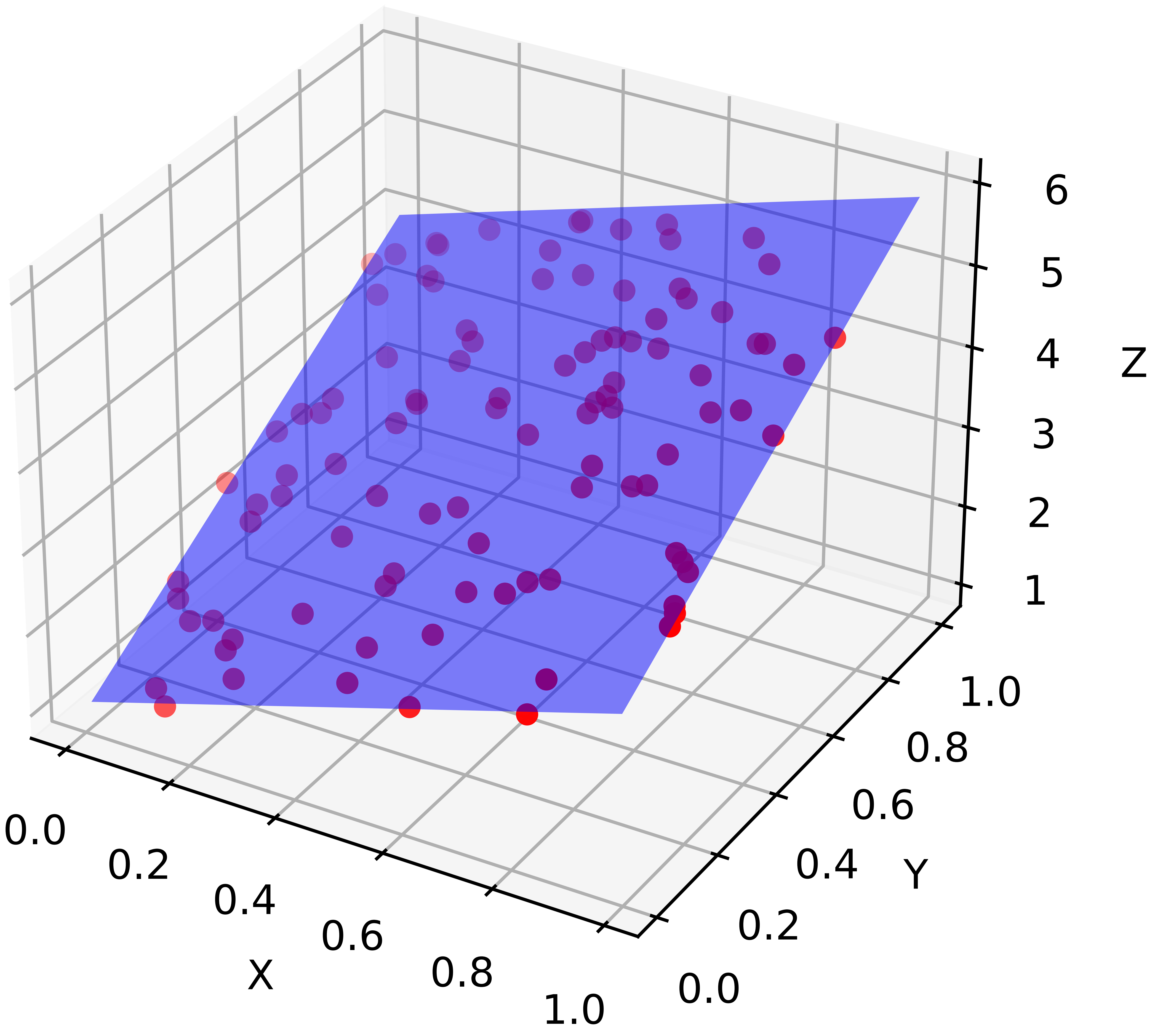
3. How do we find the best hypothesis? [**Optimization**]

$$\mathbf{w}_{\text{MSE}} = \text{argmin}_{\mathbf{w}} \sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x}_i)^2 = \text{argmin}_{\mathbf{w}} ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2$$

where $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$ $\mathbf{X} = \begin{bmatrix} \leftarrow \mathbf{x}_1^T \rightarrow \\ \leftarrow \mathbf{x}_2^T \rightarrow \\ \vdots \\ \leftarrow \mathbf{x}_N^T \rightarrow \end{bmatrix}$ $\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_N \end{bmatrix}$

Linear Regression: Linear Functions

Fitting a plane
to 2D points



Linear Regression

Consider a set of predictor (*independent*) variables $x_1, \ldots, x_d$ corresponding to an outcome (*dependent*) variable y. Regression is the problem of estimating y as a function of $x_1, \ldots, x_d$. In ***Linear Regression***, the relationship between y and $x_1, \ldots, x_d$ uses a linear model, that is it is linear in its parameters:

$$y = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d + \epsilon$$