Foundations of Machine Learning (CS 725)
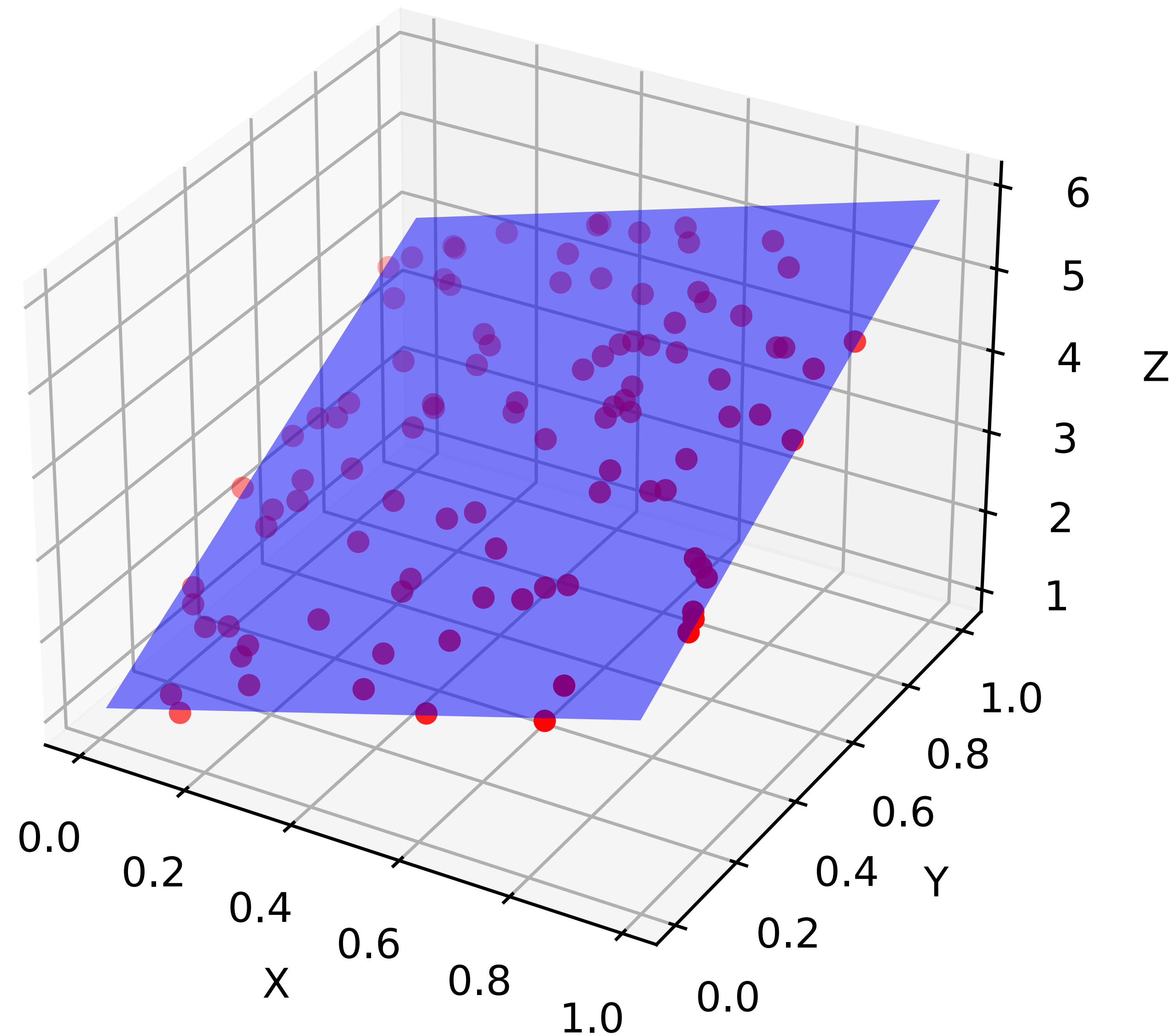## FALL 2024

Lecture 3:
- Basis Functions and Gradient Descent
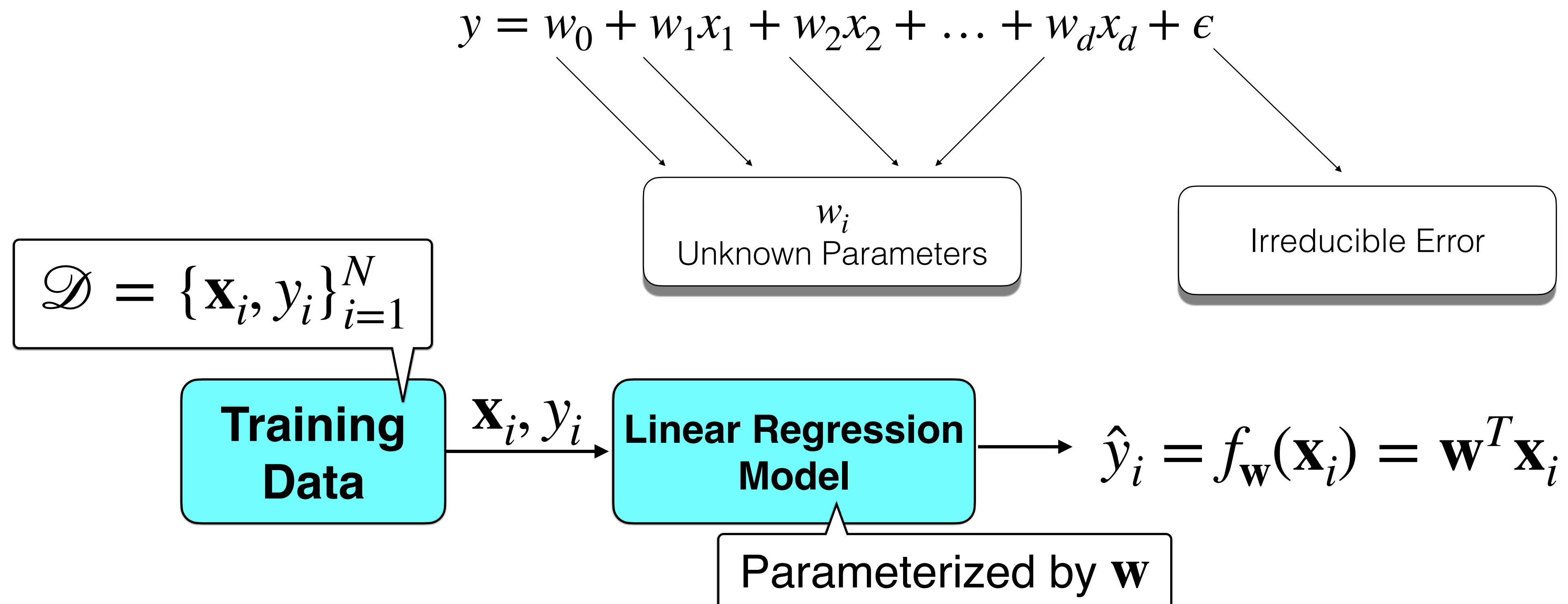
Instructor: Preethi Jyothi

Linear Regression: Linear Functions

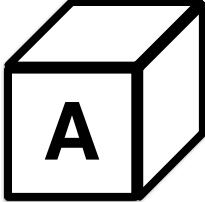Fitting a plane to 3D points in $\mathbb{R}^3, (x, f(x))$
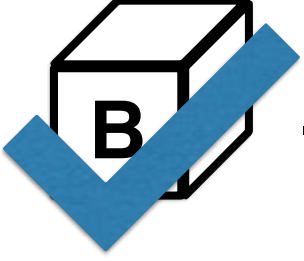
Linear Regression

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:

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

w_i
Unknown Parameters

Irreducible Error

$$\mathscr{D} = \{\mathbf{x}_i, y_i\}_{i=1}^{N}$$

Training Data

$\mathbf{x}_i, y_i$

Linear Regression Model

$$\hat{y}_i = f_{\mathbf{w}}(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i$$
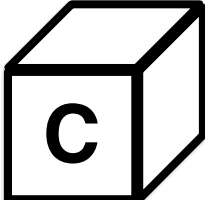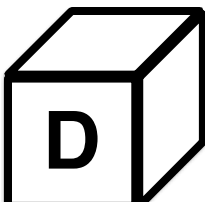
Parameterized by $\mathbf{w}$

Question

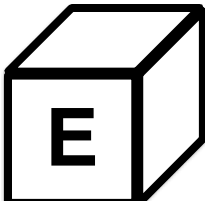Let $\mathbf{w}$ be the solution you obtain by optimizing the standard least-squares objective for linear regression. What solution would you obtain if each of the input features (not the labels) are multiplied by a factor of 25 before estimating the regression weights? Pick the correct option.

A. $25\mathbf{w}$

B. ✓ $\dfrac{\mathbf{w}}{25}$

C. $5\mathbf{w}$

D. $\dfrac{\mathbf{w}}{5}$

E. $\mathbf{w}$

Linear Regression: Overview of the Pipeline

$$\mathscr{D} = \{\mathbf{x}_i, y_i\}_{i=1}^{N}$$

Training Data $\xrightarrow{\mathbf{x}_i, y_i}$ **Model** $\longrightarrow$ $\hat{y}_i = f_{\mathbf{w}}(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i$

Parameterized by $\mathbf{w}$

Training Data $\xrightarrow{\mathbf{x}_i}$ **Feature Extraction** $\xrightarrow{\mathbf{\Phi}(\mathbf{x}_i)}$ **Model** $\longrightarrow$ $\hat{y}_i = f_{\mathbf{w}}(\mathbf{x}_i) = \mathbf{w}^T\mathbf{\Phi}(\mathbf{x}_i)$

Linear Regression With Basis Functions

$$\mathbf{\Phi}(\mathbf{x}_i) = \begin{bmatrix} \phi_0(\mathbf{x}_i) \\ \phi_1(\mathbf{x}_i) \\ \vdots \\ \phi_k(\mathbf{x}_i) \end{bmatrix}$$