

# Explanatory Notes for 6.390

Shauntclair Ruiz (Current TA)

Fall 2022

## Multiple Data Points in a Matrix

### Summing over data points

Currently, when using our **objective** function, we have to **sum** over **every** single data point. For the 1D case, this means we have to do:

$$J = \frac{1}{n} \sum_{i=1}^n (\theta x^{(i)} - y^{(i)})^2 \quad (1)$$

This is a bit of a hassle - it **forces** us to use  $x^{(i)}$  notation, and we have to be conscious of that **sum**.

By using **vectors** above, we were able to work with **many** variables  $\theta_k$  at the same time, making it easier to **represent** and **work** with them in the future.

Can we do the **same** here - combining many **data points** into one object, rather than many **variables**?

### Summing with Vectors: Row Vectors

We want to represent **addition** using **vectors**. We did that when we were adding  $x_k \theta_k$  terms with a **dot product**.

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_d \end{bmatrix} \cdot \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix} \quad (2)$$

But, dot products also include **multiplication**. Above, our terms are **squared**. So, we can multiply  $(\theta x^{(i)} - y^{(i)})$  times itself!

$$J = \frac{1}{n} \sum_{i=1}^n (\theta x^{(i)} - y^{(i)})(\theta x^{(i)} - y^{(i)}) \quad (3)$$

We'll write  $r^{(i)} = \theta x^{(i)} - y^{(i)}$  to simplify our work.

$$J = \frac{1}{n} \sum_{i=1}^n r^{(i)} * r^{(i)} \quad (4)$$

In a dot product, we **add** the **dimensions** together. So, we'll give each term in our sum its own **dimension**.

$$J = \frac{1}{n} \sum_{i=1}^n r^{(i)} * r^{(i)} = \frac{1}{n} \begin{bmatrix} r^{(1)} \\ r^{(2)} \\ r^{(3)} \\ \vdots \\ r^{(d)} \end{bmatrix} \cdot \begin{bmatrix} r^{(1)} \\ r^{(2)} \\ r^{(3)} \\ \vdots \\ r^{(d)} \end{bmatrix} \quad (5)$$

We've got a single vector we could call  $R$ .

We could make it a **column vector**, but we already use the **rows** to indicate the **dimensions**.

$$\theta = \left. \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_d \end{bmatrix} \right\} \text{dimensions as rows...} \quad (6)$$

So, let's use **columns** instead: each **column** will be a **data point**: we'll use a **row vector** ( $1 \times n$ ).

$$R = \overbrace{\begin{bmatrix} r^{(1)} & r^{(2)} & r^{(3)} & \dots & r^{(d)} \end{bmatrix}}^{\text{data points as columns!}} \quad (7)$$

### 0.0.1 Going from $x$ to $X$

We can do the same for our input data  $x^{(i)}$ :

#### Notation 1

We can store all of our 1-D **data points** in a **row vector**:

$$X = \begin{bmatrix} x^{(1)} & x^{(2)} & x^{(3)} & \dots & x^{(d)} \end{bmatrix}$$

$$Y = \begin{bmatrix} y^{(1)} & y^{(2)} & y^{(3)} & \dots & y^{(d)} \end{bmatrix}$$

Since we had  $r^{(i)} = (\theta x^{(i)} - y^{(i)})$ , we can write

$$R = \theta X - Y \quad (8)$$

Still in the 1D case!

We can write our **objective function** as

$$J = \frac{1}{n} \begin{bmatrix} \mathbf{r}^{(1)} & \mathbf{r}^{(2)} & \mathbf{r}^{(3)} & \dots & \mathbf{r}^{(d)} \end{bmatrix} \begin{bmatrix} \mathbf{r}^{(1)} \\ \mathbf{r}^{(2)} \\ \mathbf{r}^{(3)} \\ \vdots \\ \mathbf{r}^{(d)} \end{bmatrix} \quad (9)$$

So, we can write compactly:

$$J = \frac{1}{n} \mathbf{R} \mathbf{R}^T \quad (10)$$

Let's expand this back out with  $\mathbf{R} = \theta \mathbf{X} - \mathbf{Y}$ :

### Concept 2

In 1-D, we can use row vectors to sum our data points as

$$J = \frac{1}{n} (\theta \mathbf{X} - \mathbf{Y})(\theta \mathbf{X} - \mathbf{Y})^T$$

We've successfully **removed** the **sum**!

This format **stores** all of our **data points in one object**, just like how we wanted.