

# DS-DC-13

# LECTURE NOTES

*Jim Simpson*  
*Data Scientist, Sotera*

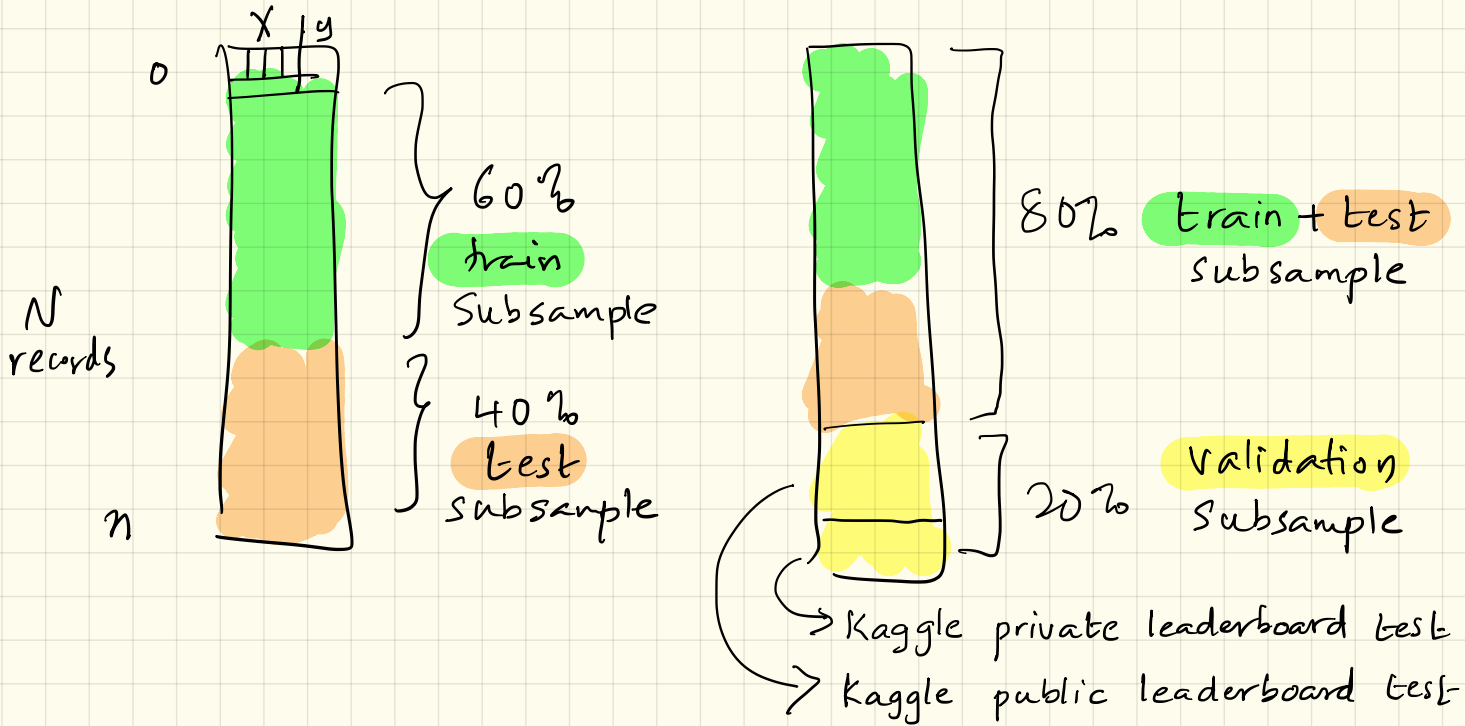
LESSON 07

# LESSON 7

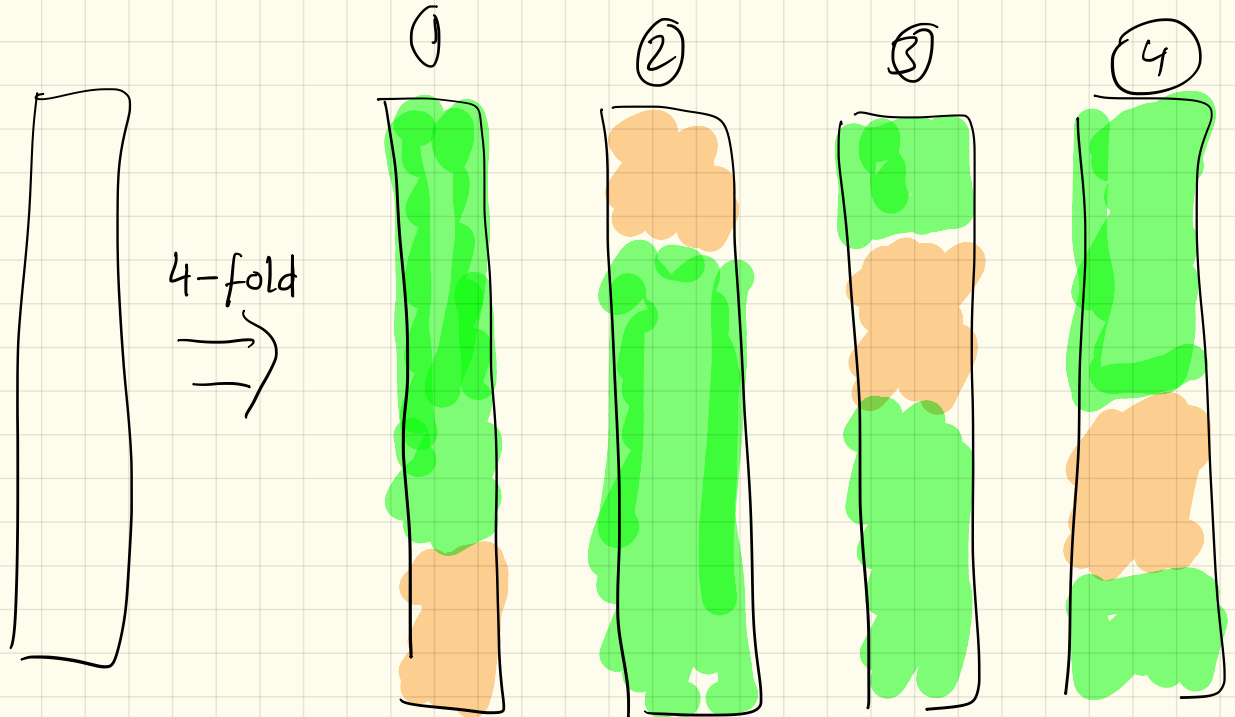
## LEARNING OBJECTIVES

- LOSS FUNCTIONS
  - MEAN SQUARED ERROR
- CROSS VALIDATION
  - K-FOLDS VALIDATION
- REGULARIZATION
  - PREVENT OVERFITTING
- GRID SEARCH
  - FOR HYPERPARAMETER TUNING
- GRADIENT DESCENT
  - FOR MODEL WEIGHT LEARNING

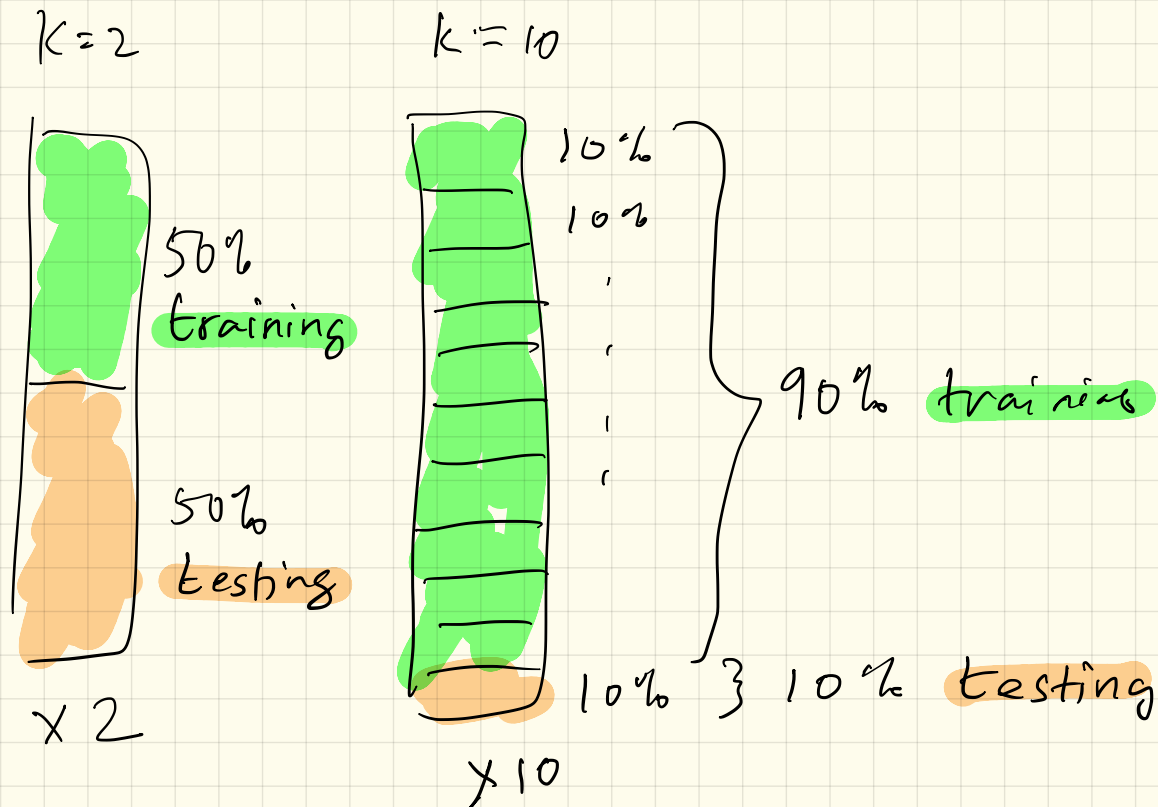
# CROSS VALIDATION



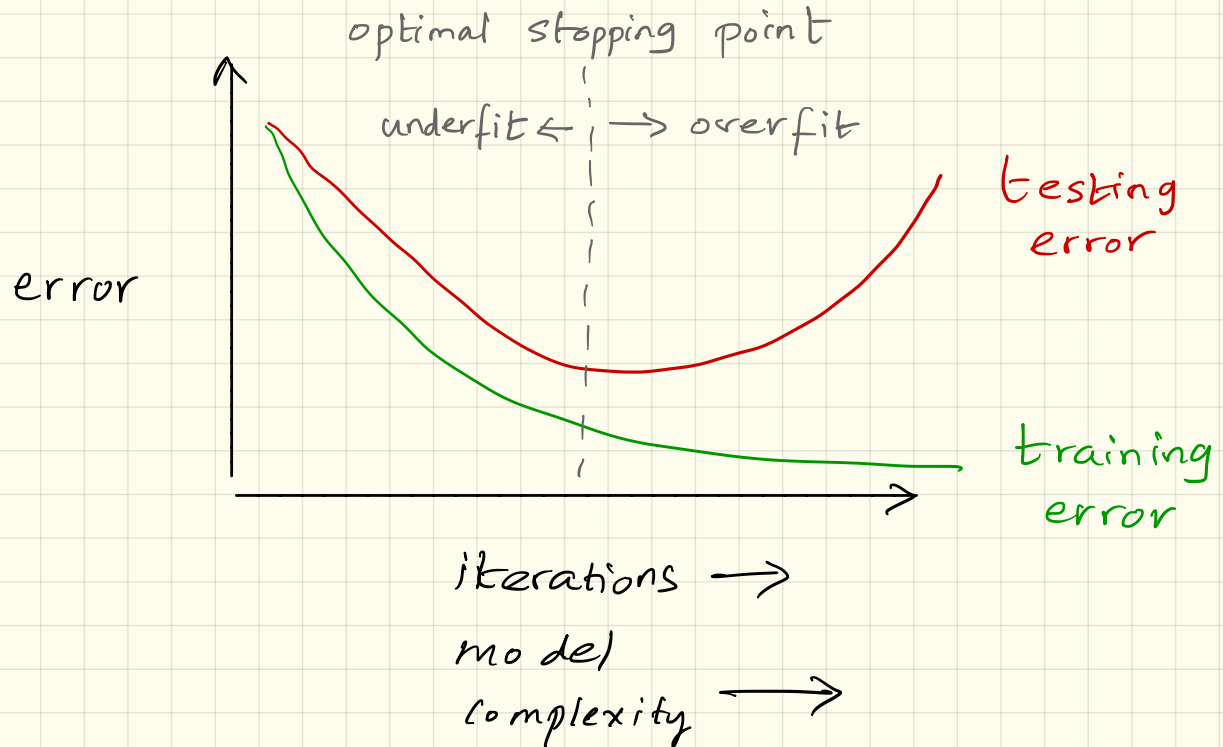
# K-FOLDS CROSS VALIDATION



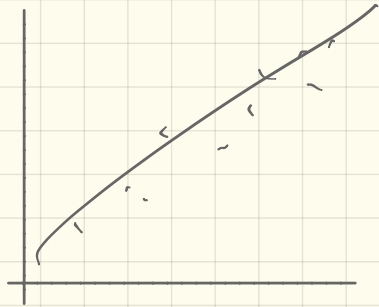
# 2-FOLDS vs. 10-FOLDS



# USING CROSS VALIDATION TO CHECK FOR OVERFITTING



# LINEAR REGRESSION WITH REGULARIZATION TO PREVENT OVERFITTING



$$y = mx + c + R(m)$$

$$y = \beta_0 x + \beta_1 + R(\beta)$$

$$y = \theta_0 x + \theta_1 + R(\theta)$$

$$\text{weights} = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

$$R(\theta) = \alpha \cdot \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \quad \text{Lasso}$$

$$R(\theta) = \alpha \begin{bmatrix} \theta_0^2 \\ \theta_1^2 \end{bmatrix} \quad \text{Ridge}$$

# GRADIENT DESCENT

Hypothesis/  
Prediction

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Weights/  
Parameters

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

Use  
grid  
search  
to tune

Gradient  
Descent

Hyperparameters

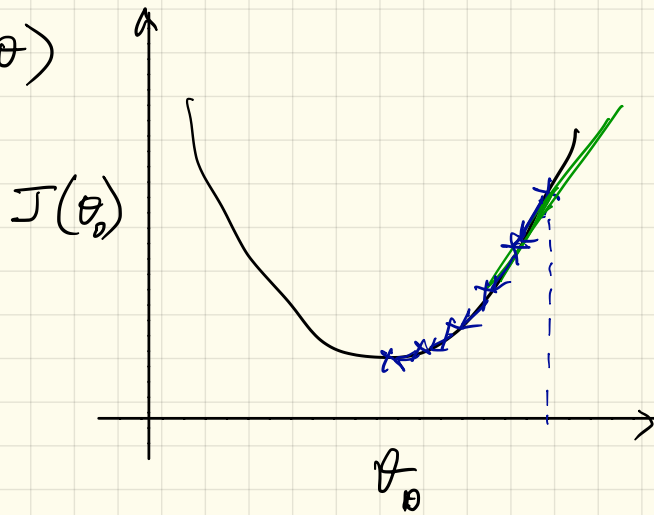
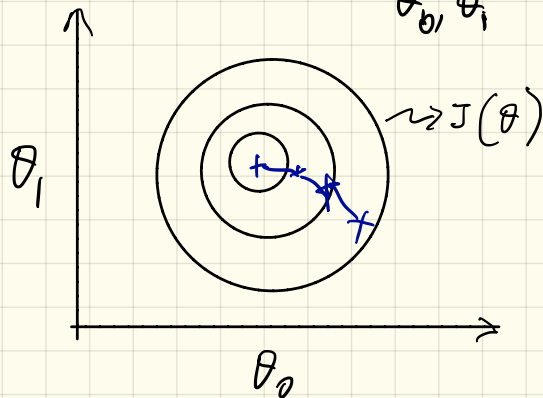
$\alpha$  learning rate  
iter iterations

Error/Objective  
Function

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \leftarrow \text{MSE}$$

Goal:

minimize  $J(\theta)$   
 $\theta_0, \theta_1$





# GRADIENT DESCENT VS GRID SEARCH

Use Grid Search to find best learning rate  
for a gradient descent

Hyperparameter of Gradient Descent

Use Gradient descent to find optimal weights  
for a linear regression model.

Parameter of Linear Regression

★ Send out additional reading