

# Scoring XGBoost Models

---



**Mike West**

MACHINE LEARNING ENGINEER



# Module Overview



**Classification and regression**

**Split your dataset into train and test**

**Cross-validation**

***K*-fold cross-validation**

**Avoid overfitting using early stops**

**Model Pima Indians Diabetes dataset**

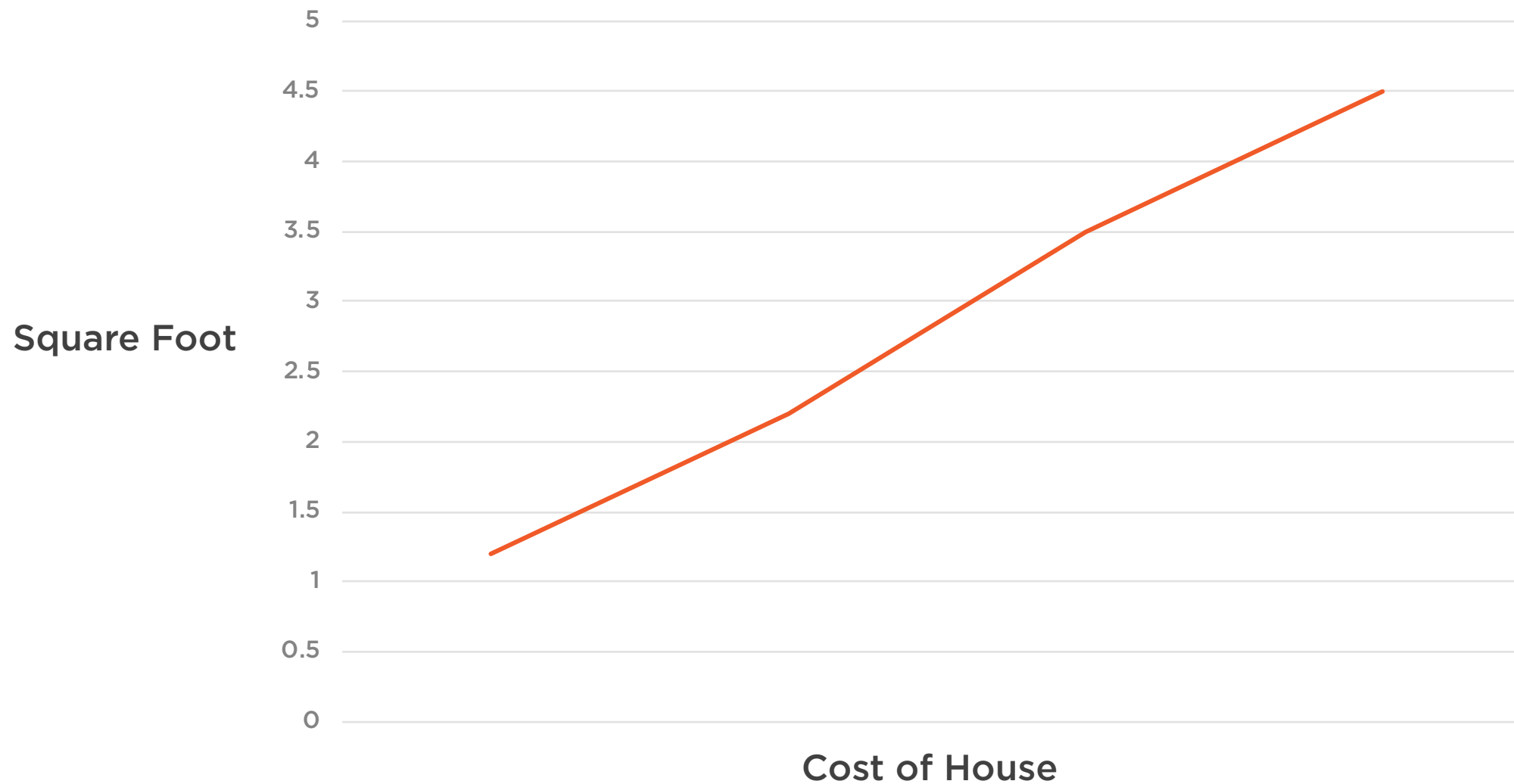


# Binary Classification

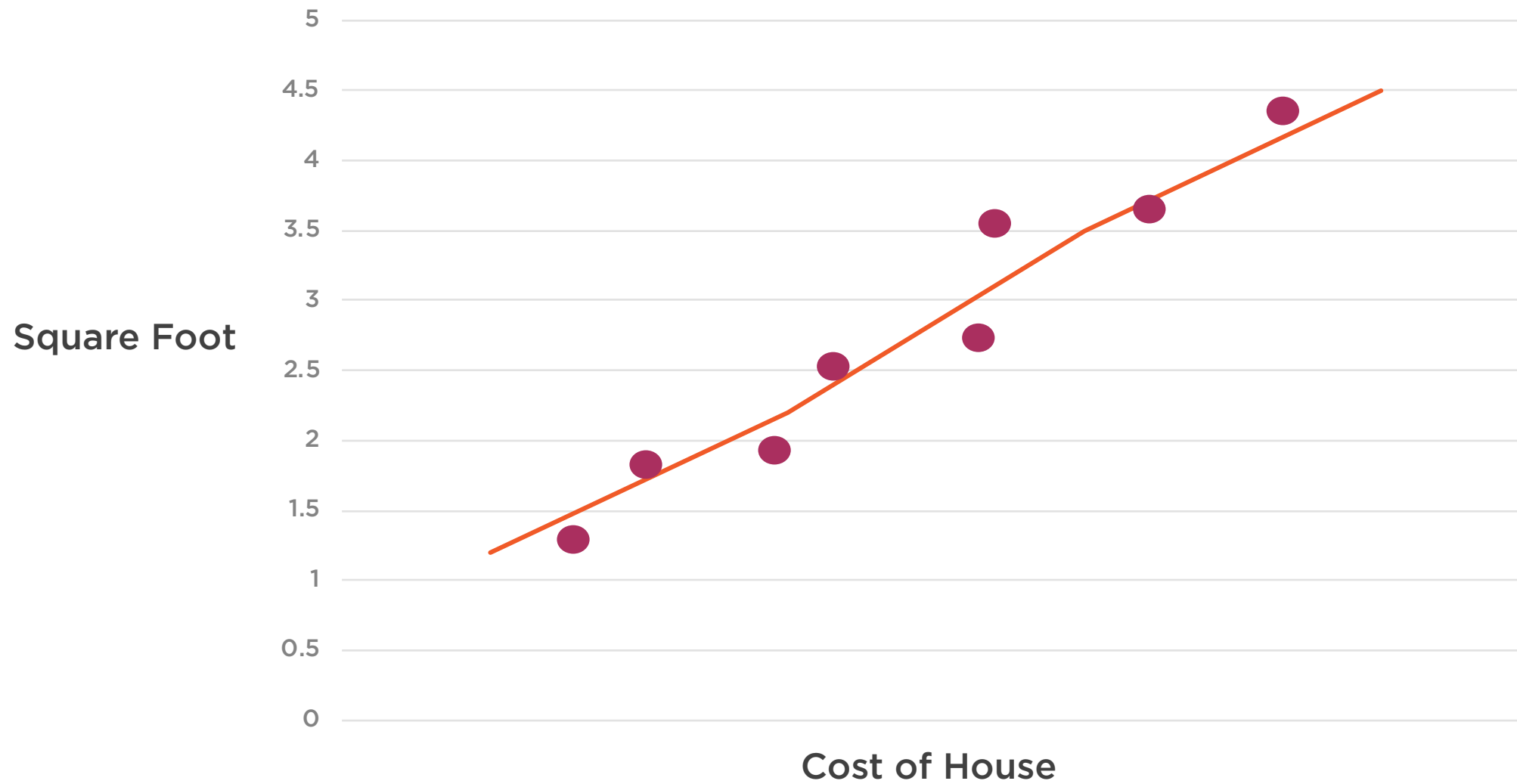
PassengerID	Survived	Name	Sex	Age
1	0	Braund, Owen	male	22
2	1	Cumings, John	female	38
3	1	Heikkinen, Laina	female	26
4	1	Futrelle, Jacques	female	35
5	0	Allen, Henry	male	35



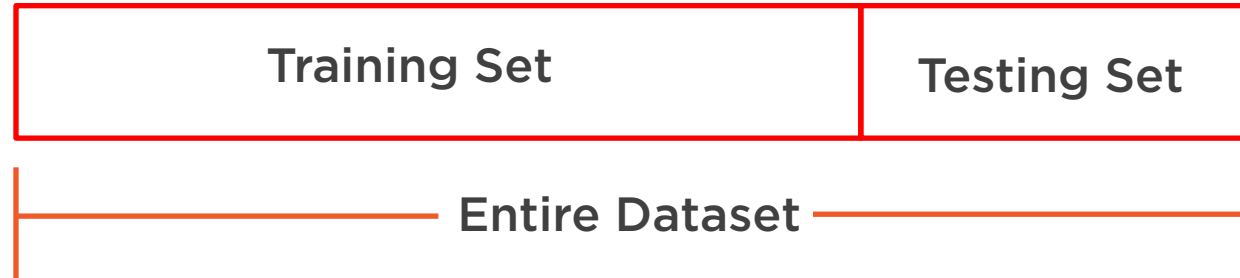
# Linear Regression



# Linear Regression



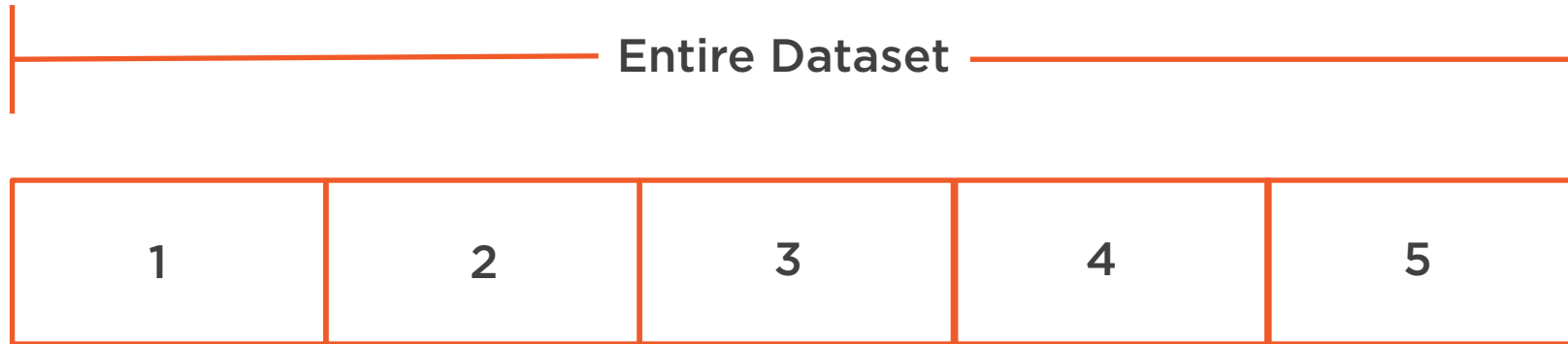
# Train / Test Split



# $K$ -fold Cross-Validation

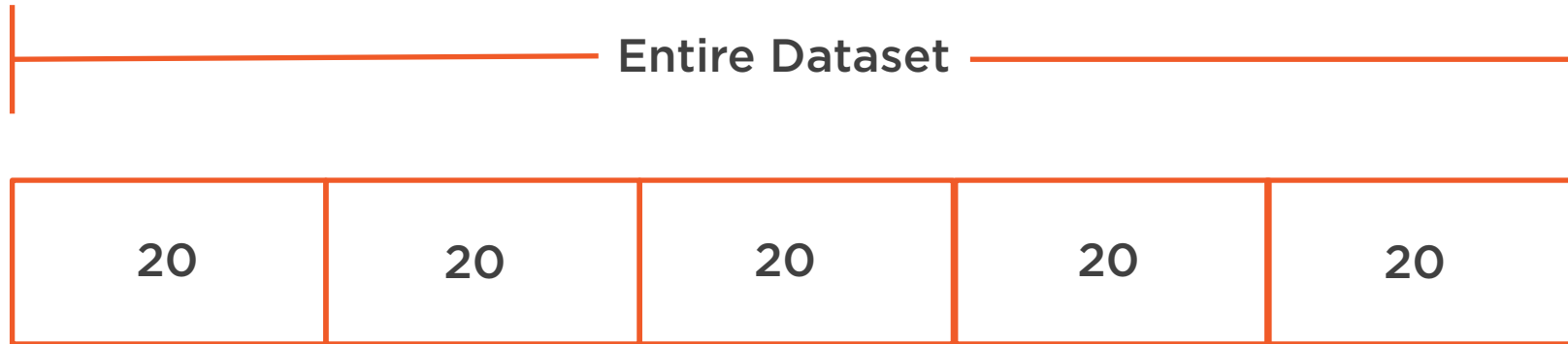


# $K$ -fold Cross-Validation





# $K$ -fold Cross-Validation



# $K$ -fold Cross-Validation



# $K$ -fold Cross-Validation



# $K$ -fold Cross-Validation

Testing				
	Testing			



# $K$ -fold Cross-Validation

Testing				
	Testing			
		Testing		



# $K$ -fold Cross-Validation

Testing				
	Testing			
		Testing		



# $K$ -fold Cross-Validation

Testing				
	Testing			
		Testing		
			Testing	



# $K$ -fold Cross-Validation

Testing				
	Testing			
		Testing		
			Testing	
				Testing





# $K$ -fold Cross-Validation

Testing					Average
	Testing				
		Testing			
			Testing		
				Testing	



# Leave-One-Out Cross-Validation



# Early Stopping



# Demo



Import your libraries

Train / test split

*K*-fold cross-validation

Early stopping

XGBoost on the dataset



# Summary



Classification and regression

Train / test split

*K*-fold cross-validation

Early stopping

Modeled the Pima Indians dataset

