**MOOC 2-MODULE 1**

[**Supervised Machine Learning: Regression**](https://www.coursera.org/learn/supervised-machine-learning-regression/home/welcome)

**I.Introduction to Machine Learning**

Introduction to Machine Learning

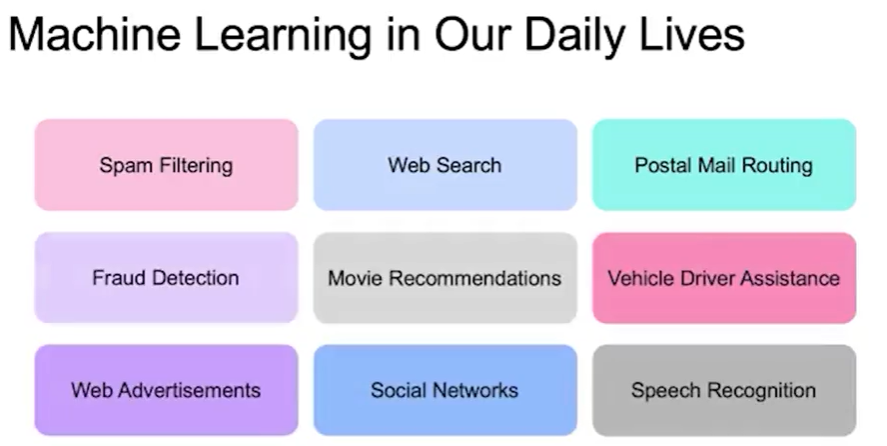
* Defines types of machine learning: supervised, unsupervised, and semi-supervised.
* Highlights the two main objectives of machine learning: predictions and interpretation.

Building Linear Regression Models

* Discusses the importance of data splitting to prevent overfitting, utilizing techniques like cross-validation.
* Introduces advanced concepts such as polynomial regression and regularization methods.

**II.Introduction to Supervised Machine Learning - Types of Machine Learning**

Understanding Machine Learning

* Machine learning allows computers to learn from data, often without knowing the underlying processes.
* Function approximation is key, enabling predictions of future values based on learned data.

Defining Artificial Intelligence and Machine Learning

* Artificial intelligence is categorized into four quadrants based on thinking vs. acting and human vs. rational behavior.
* Machine learning focuses on the thought processes, emphasizing learning as a core component.

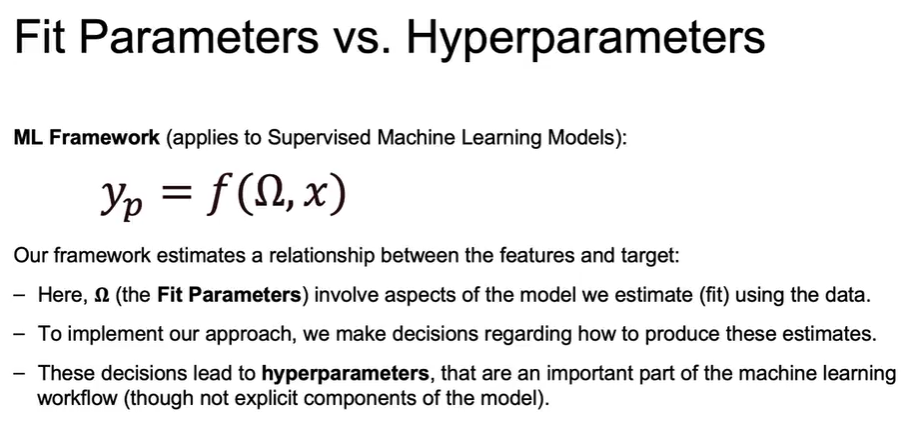
The Role of Models

* A model captures essential features of a larger reality while omitting unimportant details.
* Good models simplify complex real-world phenomena, allowing for better understanding and representation.

Applications of Machine Learning

* Machine learning is prevalent in everyday applications like spam filtering, web search ranking, route optimization, and fraud detection.
* Its integration into daily life continues to grow, impacting various sectors.

Modeling Best Practices

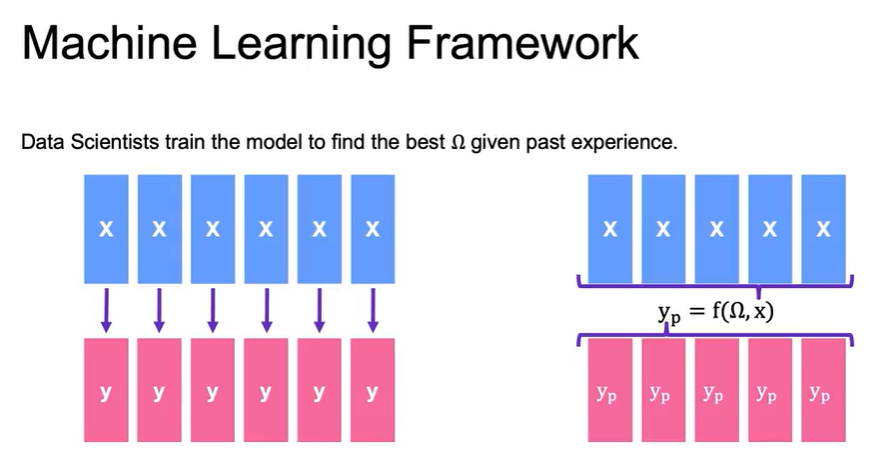
* Establish a cost function to minimize, allowing for comparison between different models.
* Develop multiple models with varying hyper-parameters to identify the best predictions.

Understanding R-squared Metric

* R-squared measures the explained variation by the model, calculated as 1 minus the unexplained variance divided by total variance.
* A higher R-squared value indicates better model performance in explaining variance.

Using Python for Regression

* Introduce the Scikit-learn library for implementing linear regression.
* Fit the model to training data and use it to make predictions on unseen test data.



**III.Supervised Machine Learning**

Interpretation

* Goal: Understand *how* variables influence outcomes.
* Models: Simple, transparent (e.g., linear regression).
* Examples:  
  + Studying customer demographics to identify sales drivers.
  + Evaluating vehicle safety features to reduce accidents.
  + Measuring marketing budget impact on movie revenue.

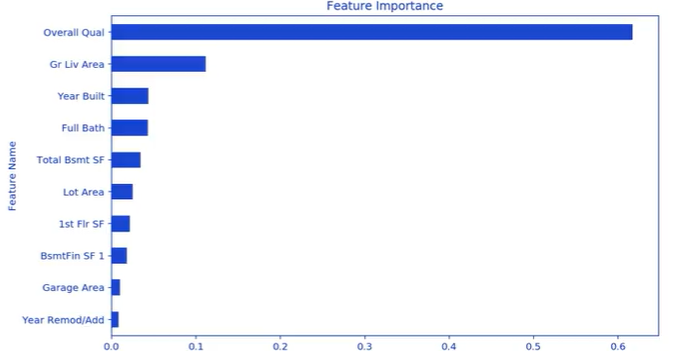
Prediction

* Goal: Maximize *accuracy* of forecasts.
* Models: Often more complex, less interpretable (e.g., random forests, neural nets).
* Examples:  
  + Predicting customer churn rates.
  + Forecasting loan defaults.
  + Predicting customer purchases from past behavior.

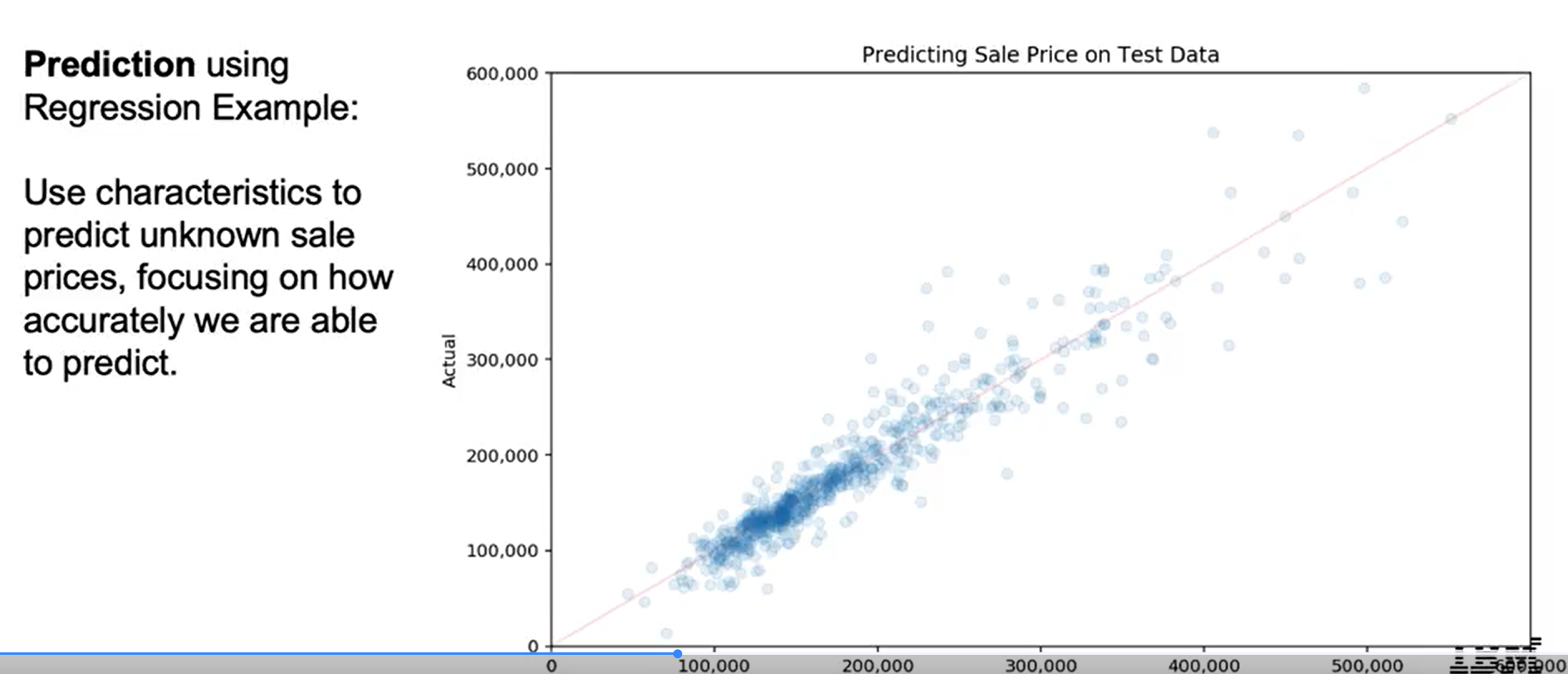
Trade-Off Considerations

* Interpretability vs. accuracy: Simple models → more explainable, Complex models → more predictive power.
* Best choice depends on business objectives (understanding vs. forecasting).

## Understanding Housing Dataset — Summary

* Target Variable: Housing price
* Features: House quality, location, number of floors, etc.
* Model Parameters: Provide insights into how each feature impacts housing prices (feature importance).  
  

## Prediction vs. Interpretation

* Prediction: Focus on how well the model forecasts housing prices.
* Interpretation: Focus on understanding the contribution of each feature to price.
* Evaluation Tool: Predicted vs. actual values plot → points closer to the diagonal = more accurate model.  
  

## Balancing Interpretation and Prediction

* Both are important in ML projects.
* Trade-off:  
  + Simple models (e.g., linear regression) → easier to interpret.
  + Complex models (e.g., deep learning) → higher predictive accuracy but less interpretable.

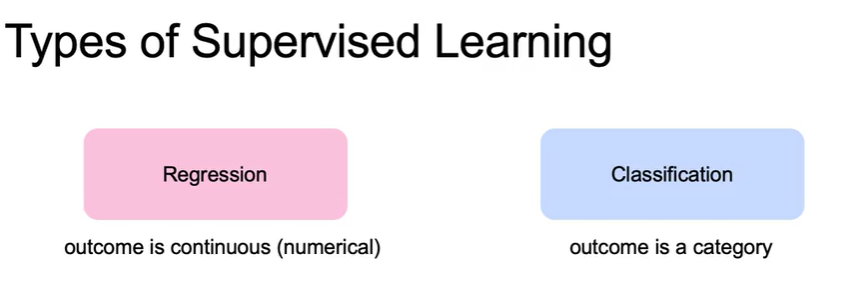
**IV.Regression and Classification Examples**

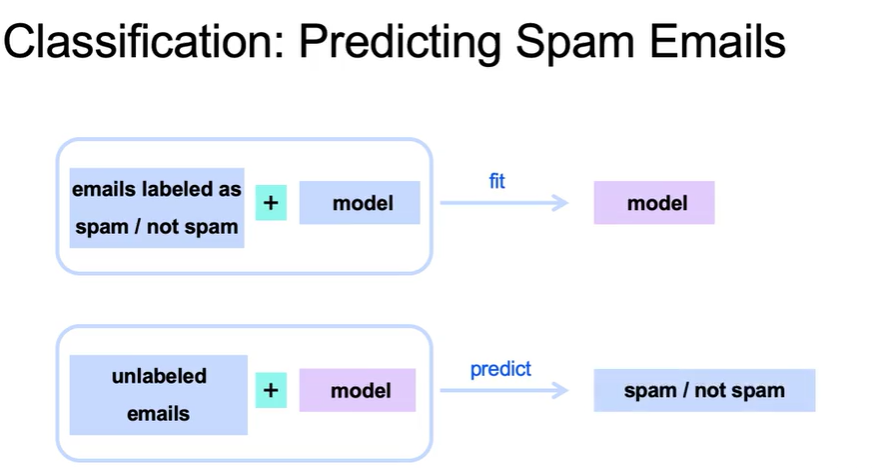
Regression

* Regression predicts continuous outcomes, such as box office revenue or housing prices.
* The process involves fitting a model to data with known outcomes to predict future values.

Classification

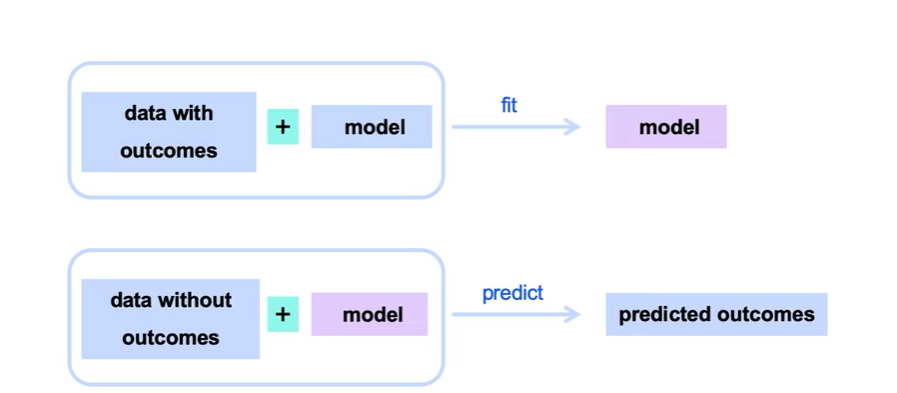
* Classification predicts categorical outcomes, like customer churn or spam detection.
* Similar to regression, a model is fitted using labeled data to classify new, unlabeled data.



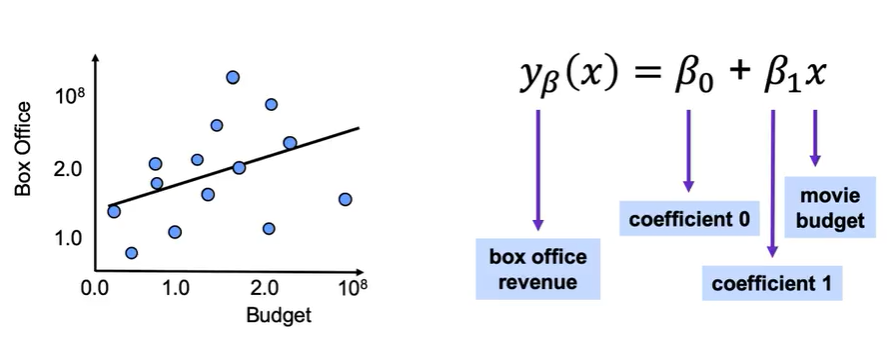


Key Concepts

* Both regression and classification require a dataset with outcomes for model training.
* The fitted models can then be used to predict outcomes for new data based on learned parameters.



**V.Introduction to Linear Regression**

* Predict box office revenue based on marketing budget.
* Linear model:  
  

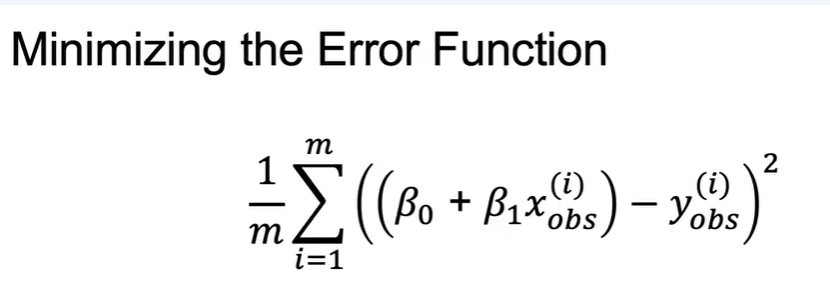


### Fitting the Model

* Find the optimal line by minimizing the cost function.
* Parameters β0,β1\beta\_0, \beta\_1β0​,β1​ are chosen to minimize the distance between predicted and actual data points.

### Error Measurement

* Error = difference between predicted and actual values.
* Common norms:  
  + L1 norm: absolute value of errors.
  + L2 norm: squared errors (Euclidean).
* Mean Squared Error (MSE) is most widely used:

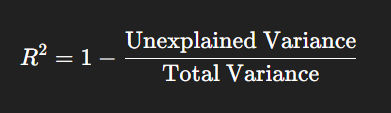
****

### Best Practices

* Define a cost function to minimize (e.g., MSE) → provides a benchmark to compare models.
* Build multiple models with different hyperparameters to identify the best performer.

### R-Squared Metric

* Formula:



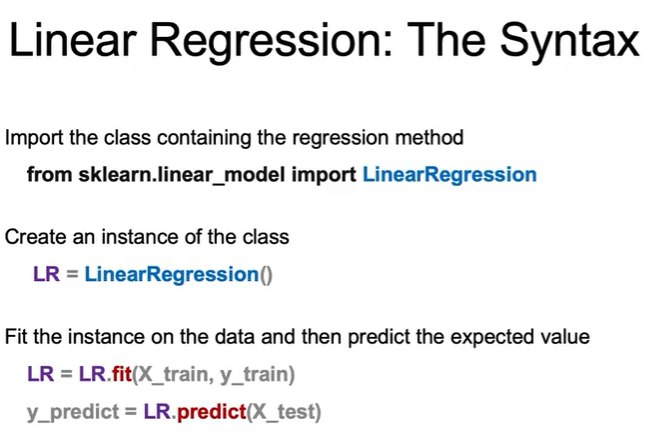
### Using Python for Regression

* Library: Scikit-learn
* Workflow:

Import and initialize linear regression model.

Fit model to training data.

Use fitted model to predict on test data.

****