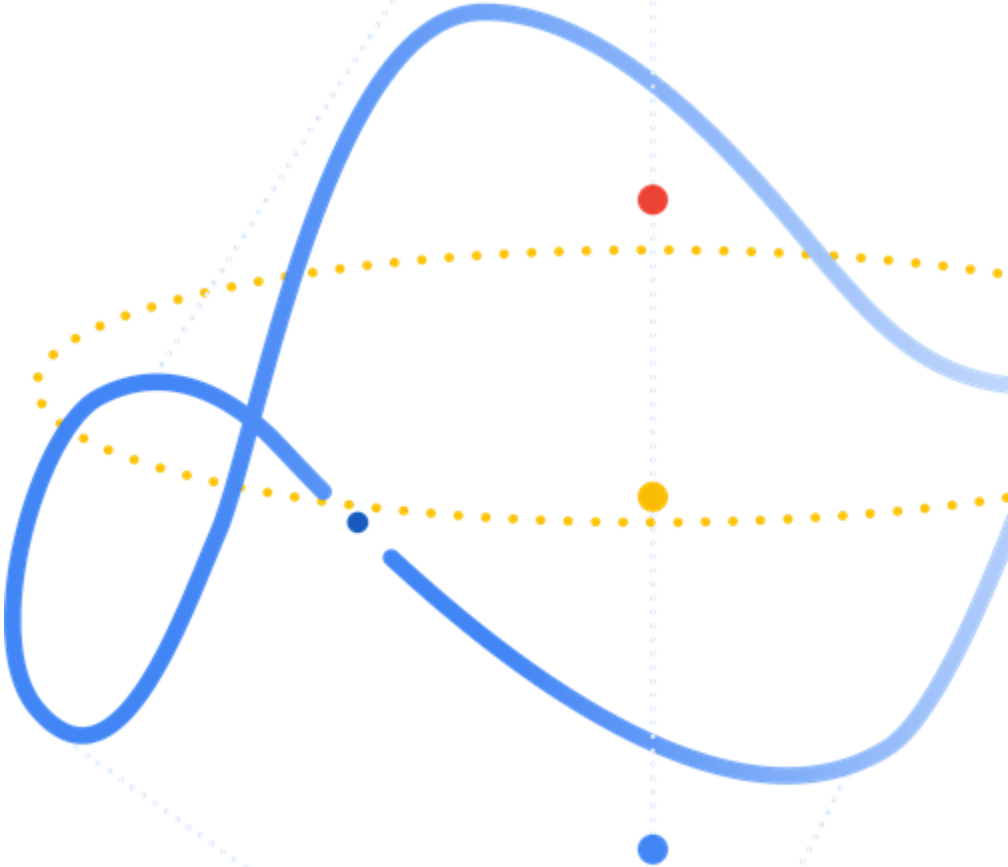
# Machine Learning



Google AI

Crowdsource
By Google

# What is Machine Learning?

Machine Learning systems take inputs (data) to make useful predictions and decisions about previously unseen pieces of data.

Machine learning is a specific field of AI where a system learns to find patterns in examples in order to make predictions.

Computers learning how to do a task without being explicitly programmed to do so.

# Machine Learning systems might:

- Label or classify data
- Predict numerical values
- Cluster similar pieces of data together
- Infer association patterns in data
- Create complex outputs

# "Machine Learning: Why or Why not?"



Read a couple of news articles involving applications of ML.

1. Would a traditional programming solution be more efficient?
1. Could a human perform the same task in less time?
1. What are the benefits of a Machine Learning model in these instances?

# Machine Learning

## Supervised

Model is trained on labeled data


stop_sign_1


stop_sign_2

stop_sign_3


stop_sign_4

## Unsupervised

Model learns patterns from unlabelled data.

# See it in action!

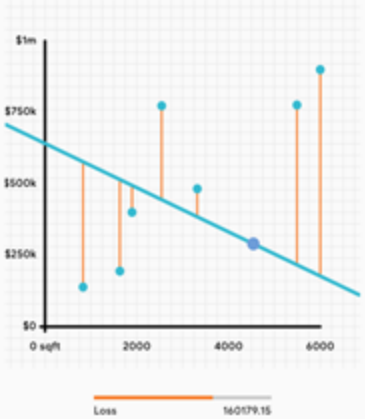
## Supervised learning

Image label verification

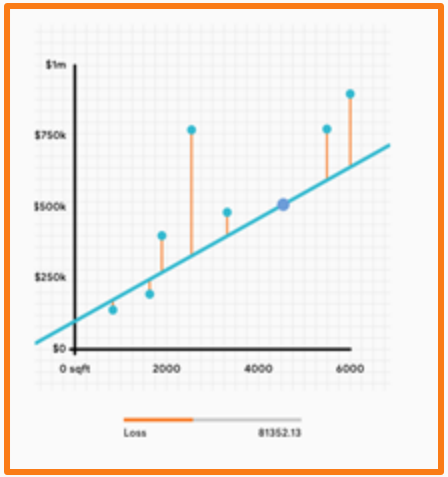
## Unsupervised learning

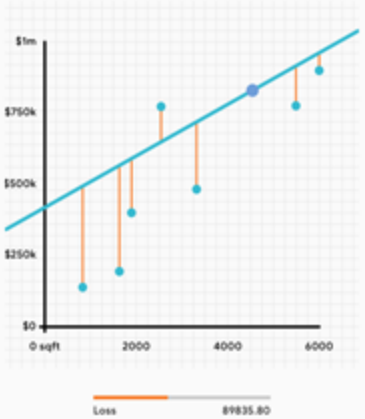
Semantic Similarity

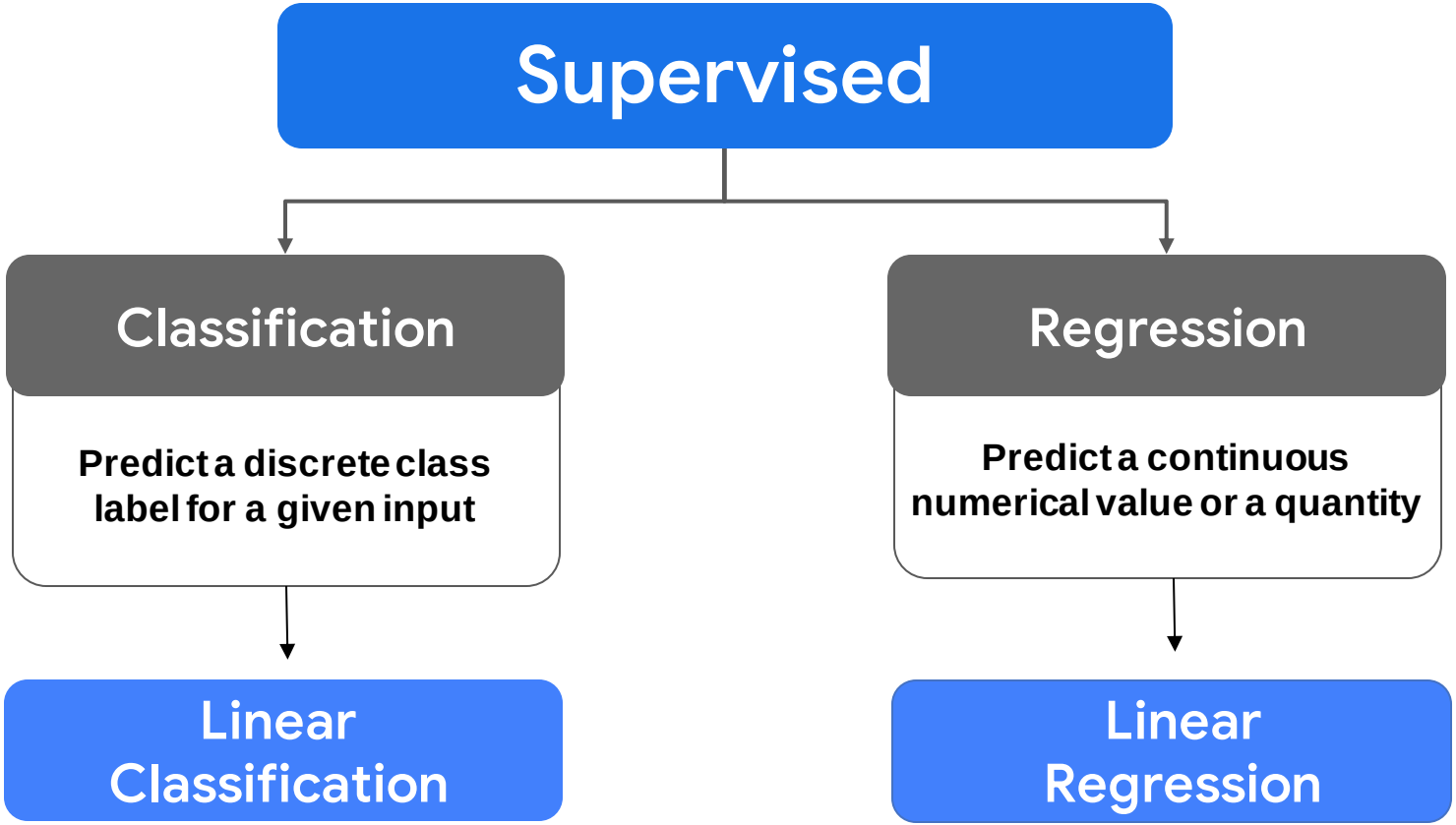
# Loss


$1m
$750k
$500k
$250k
$0
0 sqft
2000
4000
6000
Loss
160179.15


$1m
$750k
$500k
$250k
$0
0 sqft
2000
4000
6000
Loss
106367.96


$1m
$750k
$500k
$250k
$0
0 sqft
2000
4000
6000
Loss
81352.13


$1m
$750k
$500k
$250k
$0
0 sqft
2000
4000
6000
Loss
89835.80


Supervised
Classification
Predict a discrete class label for a given input
Linear Classification
Regression
Predict a continuous numerical value or a quantity
Linear Regression

# Classification

Categorizing data into predefined classes or categories based on their features.

**Process of Classification:**

- Data preparation: Collect and label dataset.
- Feature extraction: Extract relevant features.
- Training: Train classification algorithm.
- Evaluation: Assess model performance.
- Prediction: Predict class labels for new instances.

Google AI

Crowdsource
By Google

## Linear Regression

- Predicts continuous values.
- Uses linear decision boundary.
- Assumes linear relationship.
- Simple and interpretable.
- Limited for binary classification tasks.

## K-Nearest Neighbors(KNN)

- Proximity-based classification.
- Nonlinear decision boundaries.
- Requires optimal k selection.
- Simple and interpretable.
- Effective for small to medium-sized datasets.

## Random Forest

- Ensemble of decision trees.
- Reduces overfitting.
- Handles high-dimensional data.
- Provides feature importance.
- Robust to missing data and outliers.

# Regression

A statistical technique used to predict continuous numerical values based on the relationship between a dependent variable and one or more independent variables.

**Process of Regression:**

- Data preparation: Collect and preprocess dataset.
- Model selection: Choose regression algorithm.
- Training: Fit model to training data.
- Evaluation: Assess model performance.
- Prediction: Make predictions on new data instances.

Google AI

Crowdsource
By Google

## Logistic Regression

- Predicts binary outcomes.
- Estimates probabilities.
- Assumes linear relationship.
- Interpretable coefficients.
- Requires feature scaling.

## K-Nearest Neighbors(KNN)

- Uses nearby neighbors for prediction.
- Handles nonlinearity.
- Sensitive to choice of neighbors (k).
- Simple and interpretable.
- Good for small to medium-sized datasets.

## Random Forest

- Ensemble of decision trees.
- Reduces overfitting.
- Handles high-dimensional data.
- Provides feature importance.
- Robust to missing data and outliers.