Introduction

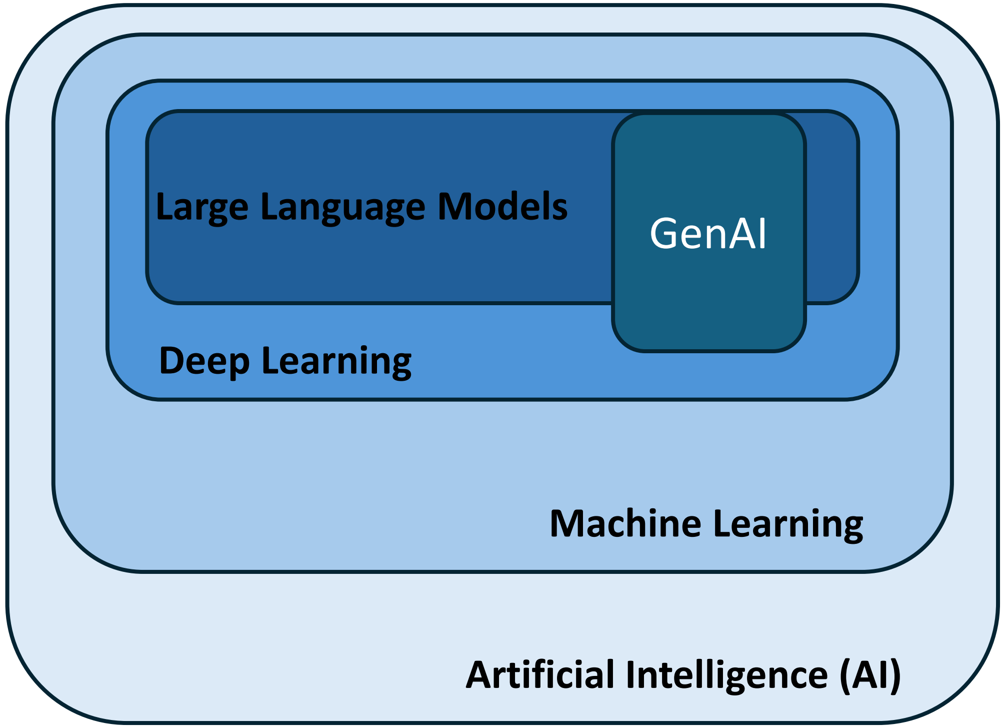
CS 335: Introduction to Large Language Models

Abdul Samad, Fisal Alvi Habib University

Contents

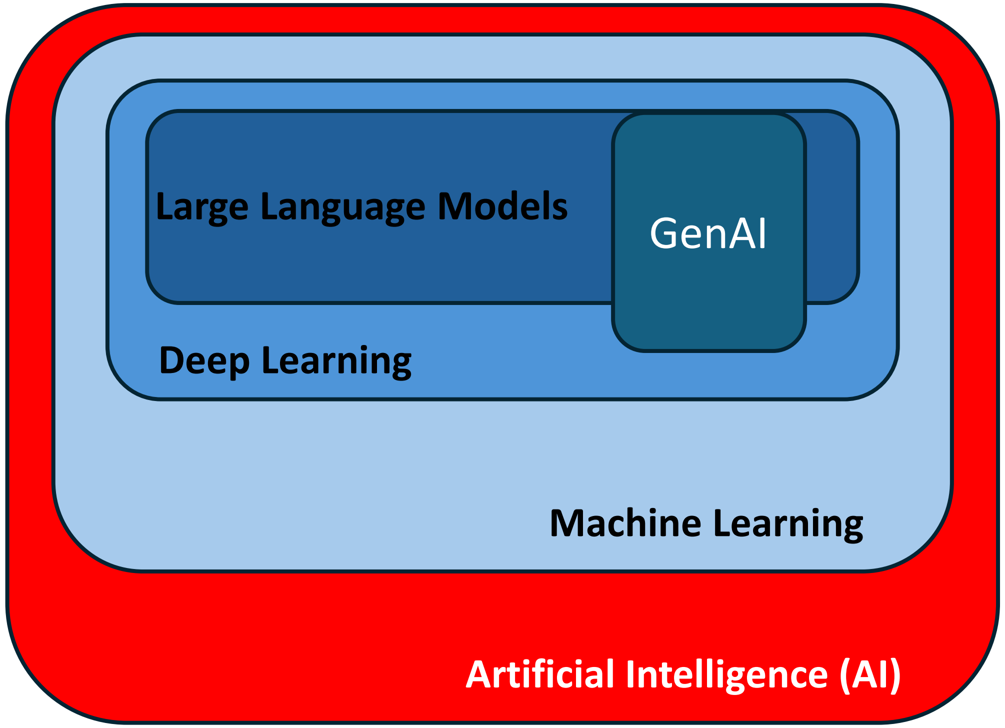
Artificial Intelligence

- Artificial Intelligence and subfields



Artificial Intelligence

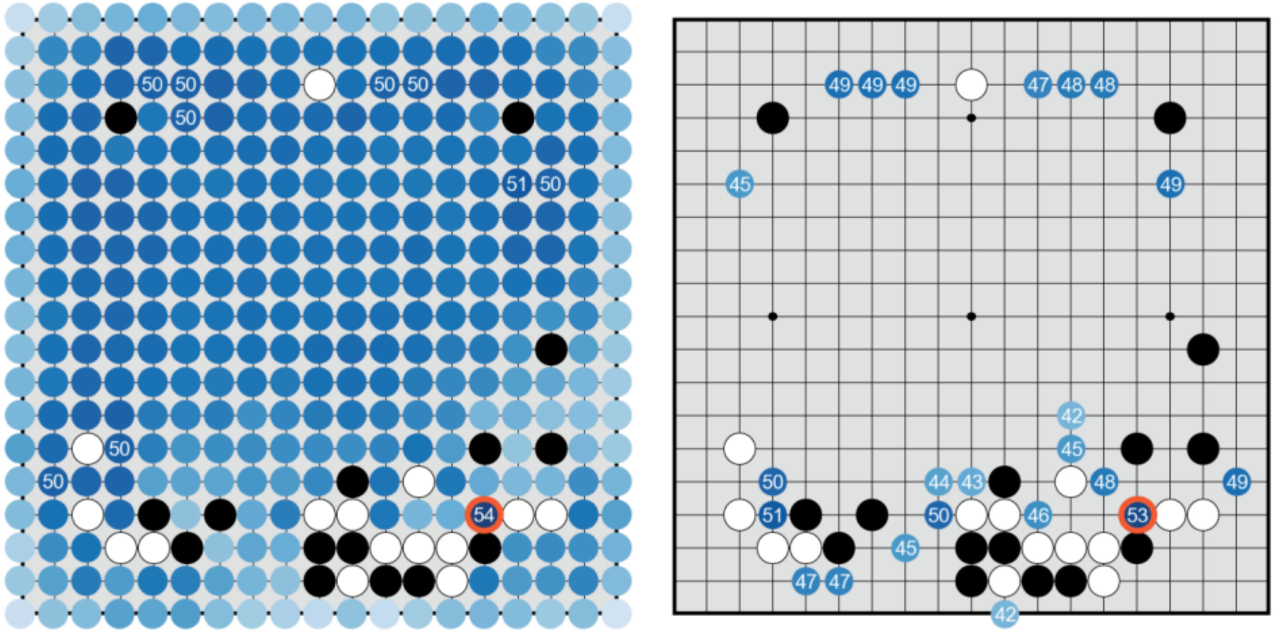
- Artificial Intelligence (AI) refers to computer programs or systems that can perform tasks that usually require human intelligence.



Artificial Intelligence

- Artificial Intelligence (AI) refers to computer programs or systems that can perform tasks that usually require human intelligence.

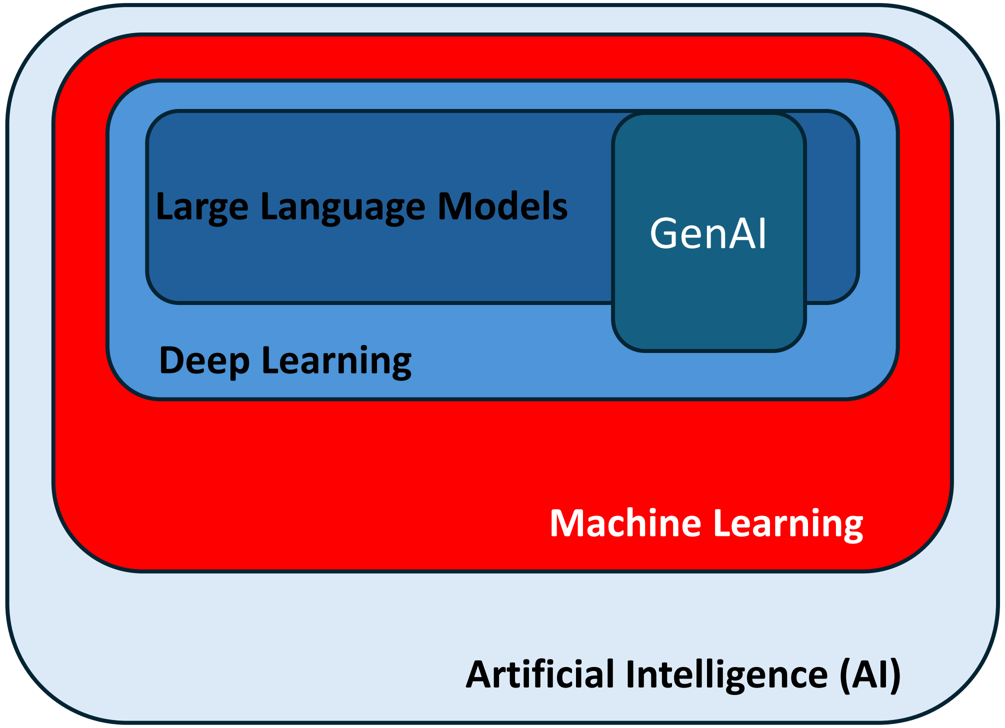
Self driving cars



AI plays Chinese game of "GO"

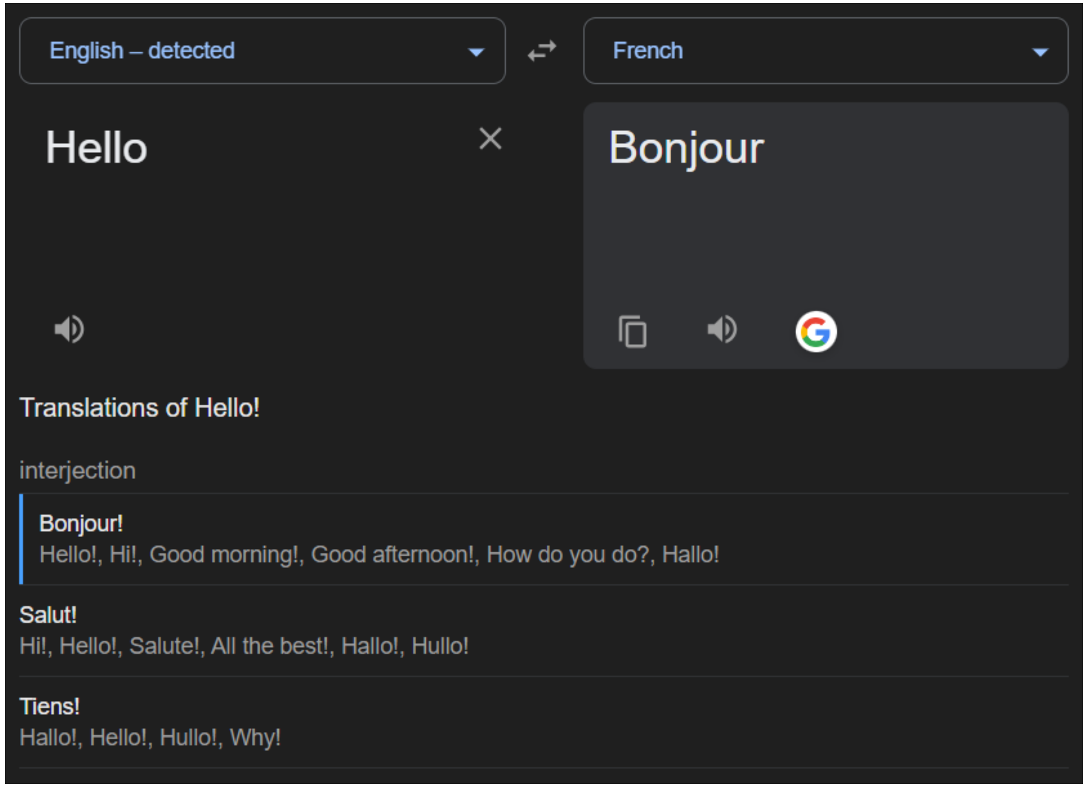
Machine Learning

- Machine Learning (ML) is a subset of AI algorithms that learn rules/patterns automatically from data.



Machine Learning

- Machine Learning (ML) is a subset of AI algorithms that learn rules automatically from data.



Machine Translation
Translate from one language to another

Clustering
Group unlabeled data into categories

Types of Machine Learning

Supervised Learning:

A type of machine learning where the model is trained on labeled data.

Examples: Image classification, spam detection, medical diagnosis, credit scoring, speech recognition.

Key Concept: Input data has corresponding output labels.

Unsupervised Learning:

A type of machine learning where the model finds patterns from unlabeled data.

Examples: Customer segmentation, anomaly detection, topic modeling, recommendation systems, clustering of genetic data.

Key Concept: No predefined labels; the algorithm identifies structure in the data.

Types of Supervised Learnings

Classification:

A supervised learning technique used to categorize data into predefined classes or labels.

Examples: Email spam detection (spam vs. not spam), disease diagnosis (positive vs. negative), image recognition (cat vs. dog).

Key Concept: Predicts discrete categories.

Regression:

A supervised learning technique used to predict continuous numerical values.

Examples: House price prediction, stock price forecasting, temperature prediction.

Key Concept: Predicts continuous values based on input features.

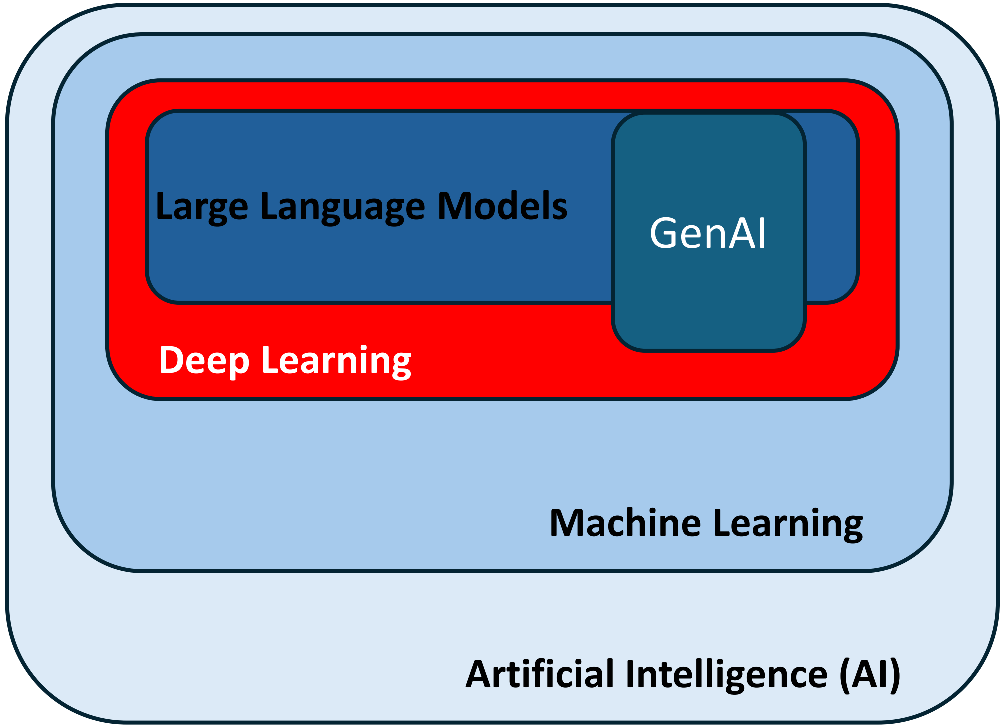
Algorithms in Machine Learning

Famous Algorithms in Machine Learning

1. Linear Regression:
2. Logistic Regression:
3. Decision Trees:
4. Support Vector Machines (SVM):
5. K-Nearest Neighbors (KNN):
6. Neural Networks:
7. Random Forest:
8. K-Means Clustering:
9. Gradient Boosting Machines (GBM):

Deep Learning

- Deep Learning (DL) is a subset of machine learning algorithms involving neural networks



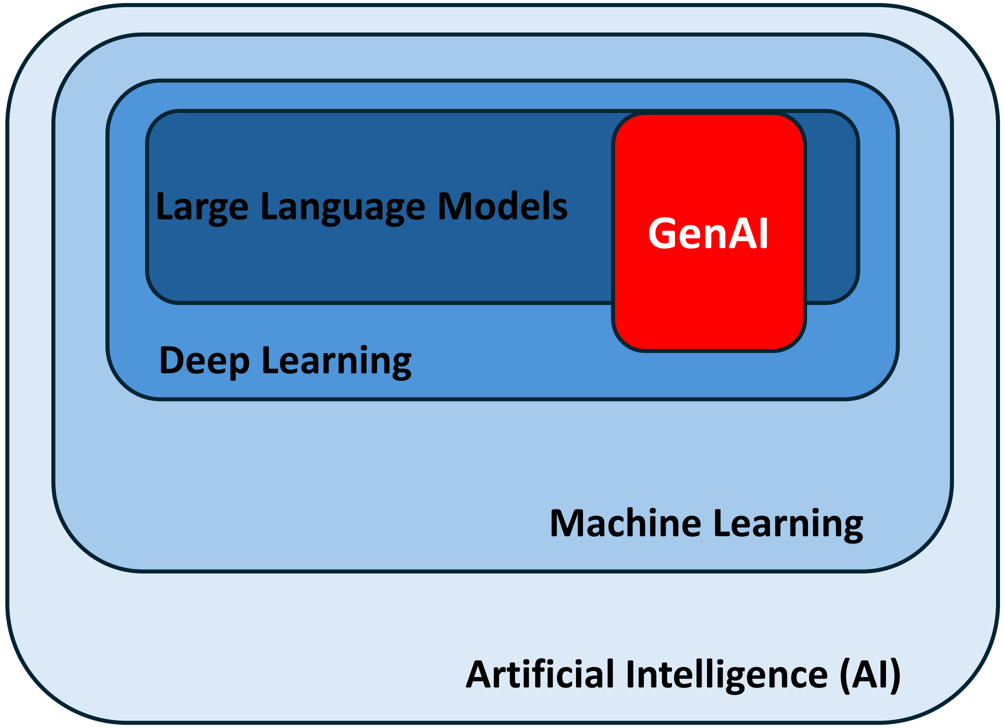
Deep Learning

- Deep Learning (DL) is a subset of machine learning algorithms involving neural networks

→ **Deep Neural Network** → **Cancerous or Not**

Generative AI

- Generative AI (GenAI) involves the use of deep neural networks to create new content, such as text, images, or various forms of media.



Generative AI

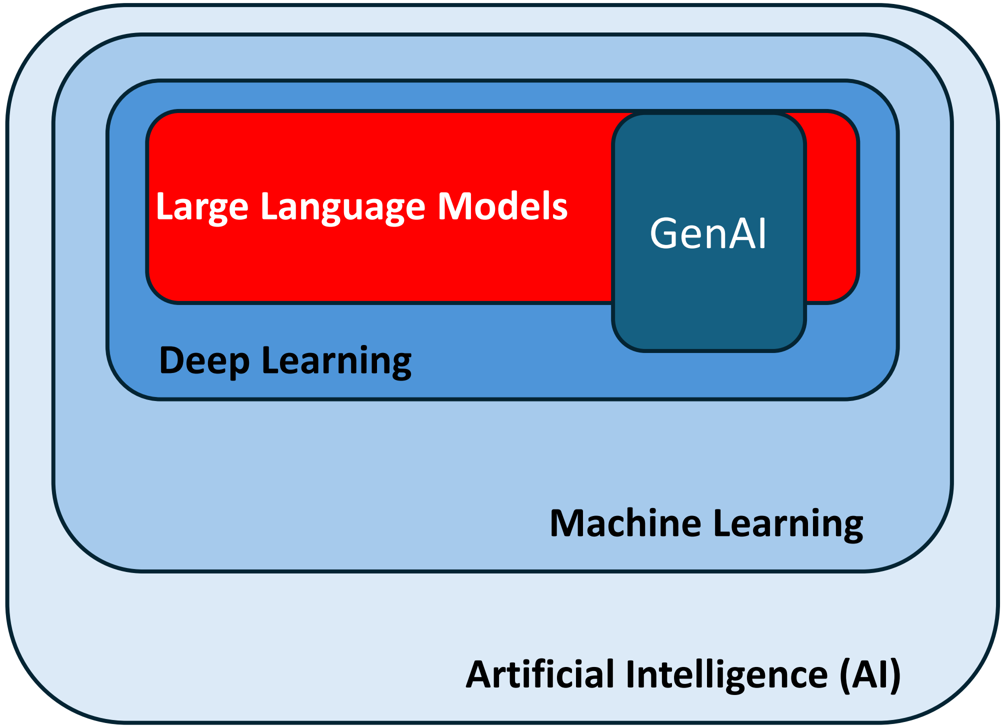
- Generative AI (GenAI) involves the use of deep neural networks to create new content, such as text, images, or various forms of media.

“Balloons tied to a cat on an empty street”

GenAI

Large Language Models

- Large Language Models (LLMs) are neural networks for parsing and generating human like text using attention mechanism.



Large Language Models

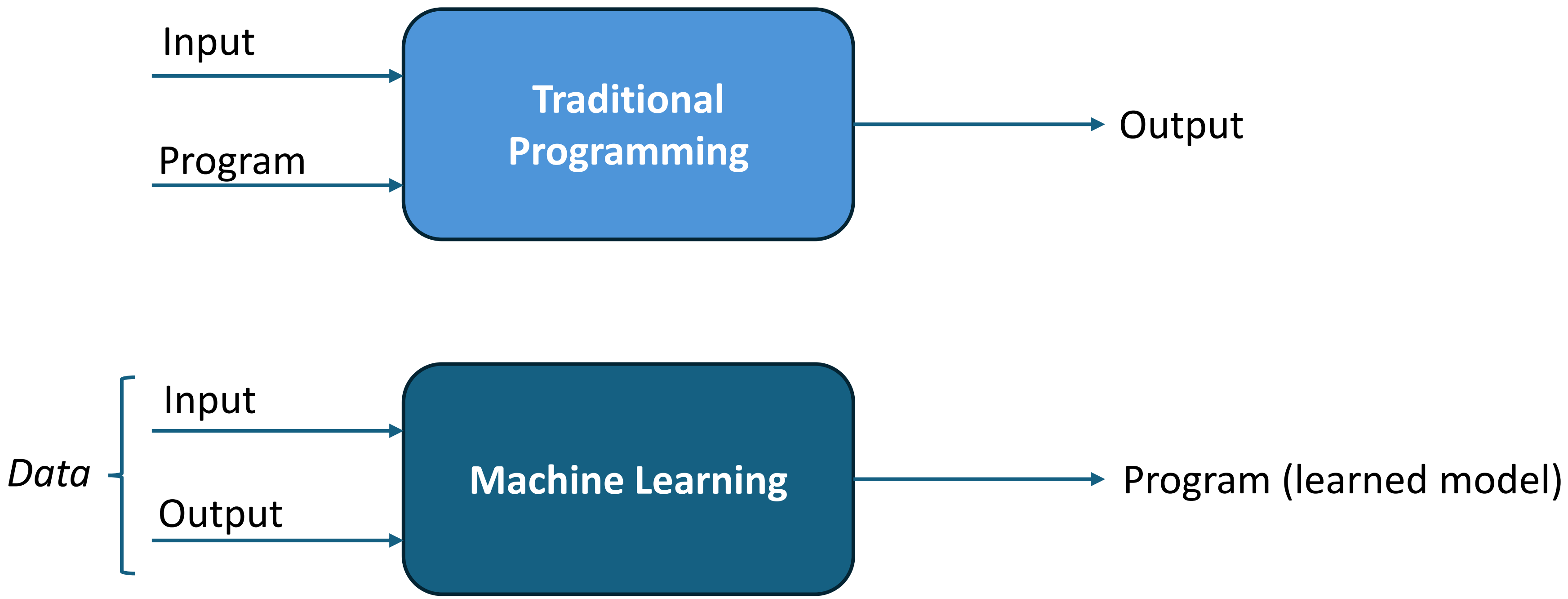
- Large Language Models (LLMs) are neural networks for parsing and generating human like text using attention mechanism.

https://www.techradar.com/computing/artificial-intelligence/best-llms

Dive into Machine Learning

What is ML?

- ML is the paradigm of approximating a function (model) from data.



Regression -House Price Prediction

Size in feet² $(\boldsymbol{x})$	**Price ($) in 1000's $(\boldsymbol{y})$**
2104	460
1416	232
1534	315
⋮	⋮

- Notation
  - x's: input variables called **features**
  - y: output variable called **target**
  - (x_i, y_i): one data point e.g. $(x_1, y_1) = (2104, 460)$
  - Data $D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots (x_n, y_n)\} = \{(x_i, y_i)\}_1^n$

Regression- House Price Prediction

Size in feet² (x)	**Price ($) in 1000's ($y$)**
2104	460
1416	232
1534	315
⋮	⋮

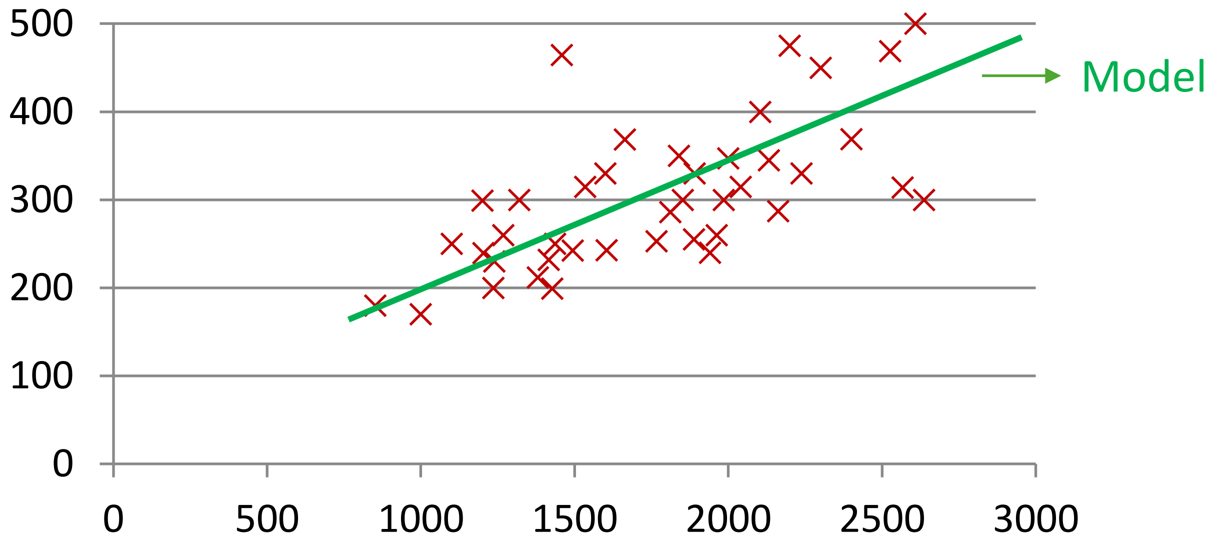
- Notation
  - x's: input variables called **features**
  - y: output variable called **target**
  - (x_i, y_i): one data point e.g. $(x_1, y_1) = (2104, 460)$
  - Data $D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots (x_n, y_n)\} = \{(x_i, y_i)\}_1^n$

Supervised learning (Regression)
both input and the corresponding correct output is available

Model

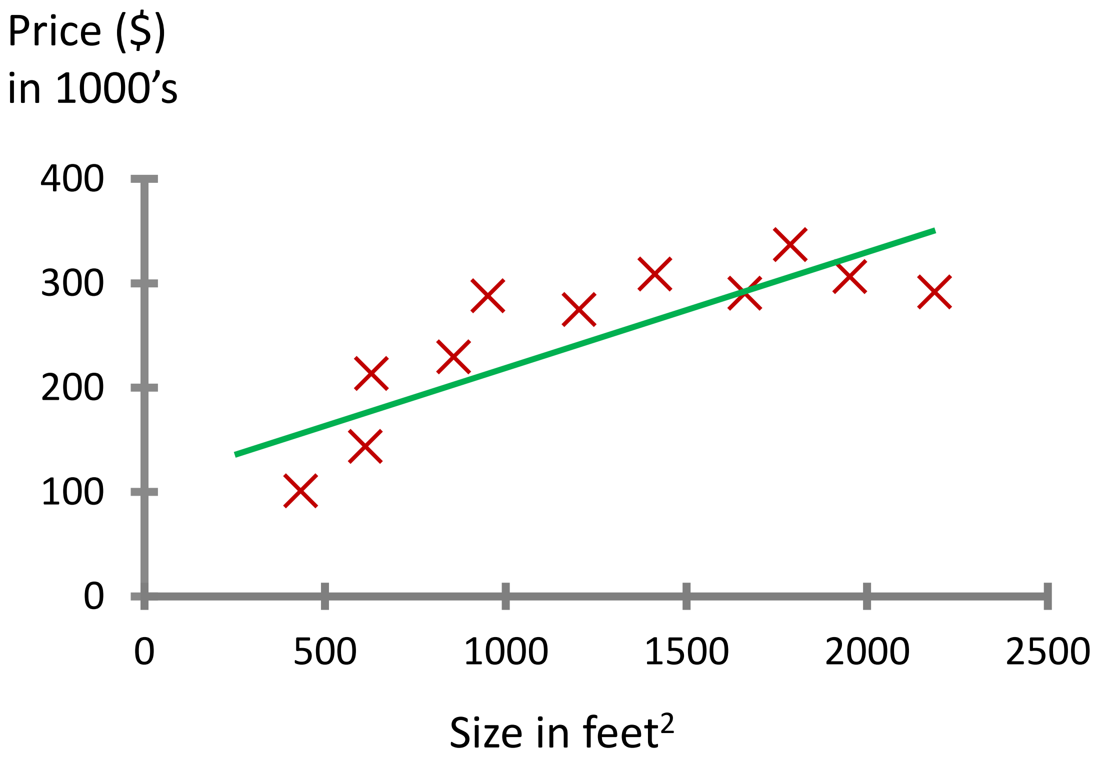
- Simplest model, fit a line to data, which describe the trend.

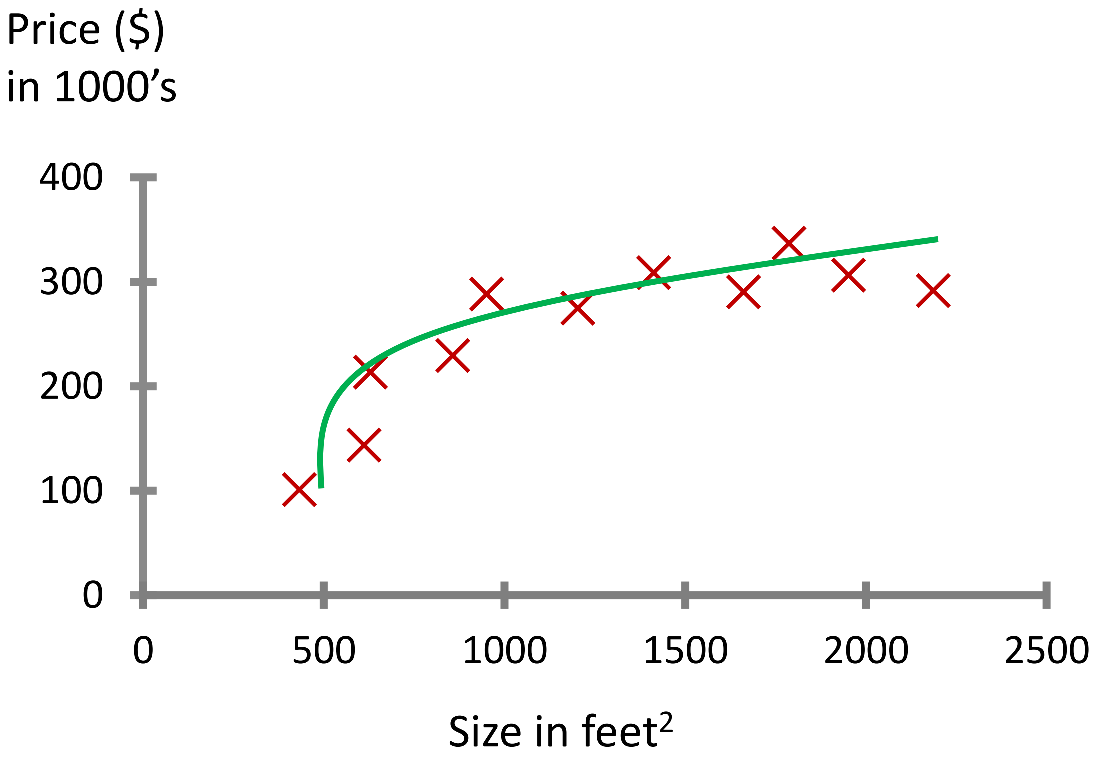
Size in feet² (x)	**Price ($) in 1000's (y)**
2104	400
1416	232
1534	315
⋮	⋮



Model

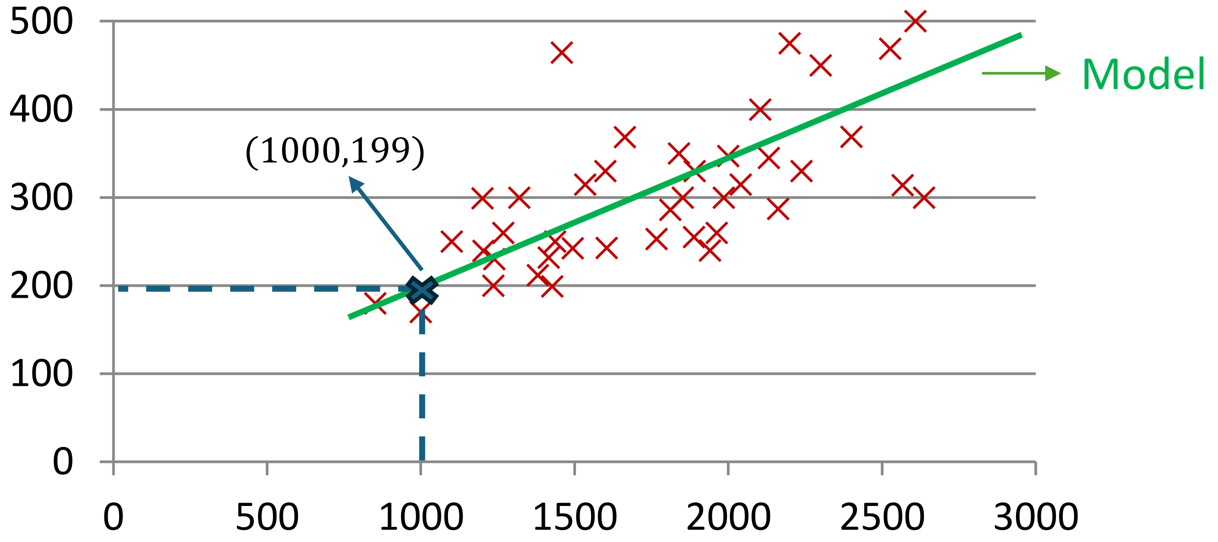
- Linear VS non-Linear Model





Prediction

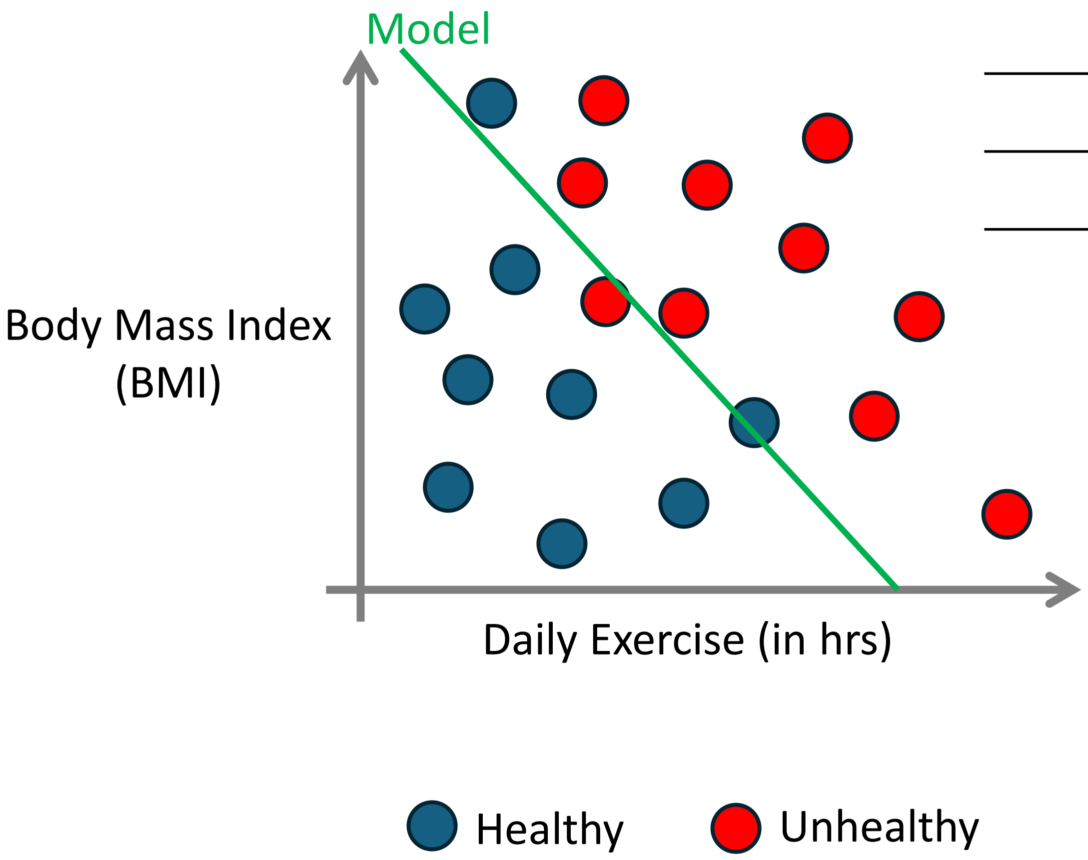
- For a new instance x use the model (line) to approximate the possible y value, called prediction (inferencing).



$x = 1000$

$\hat{y} = ?$

Classification

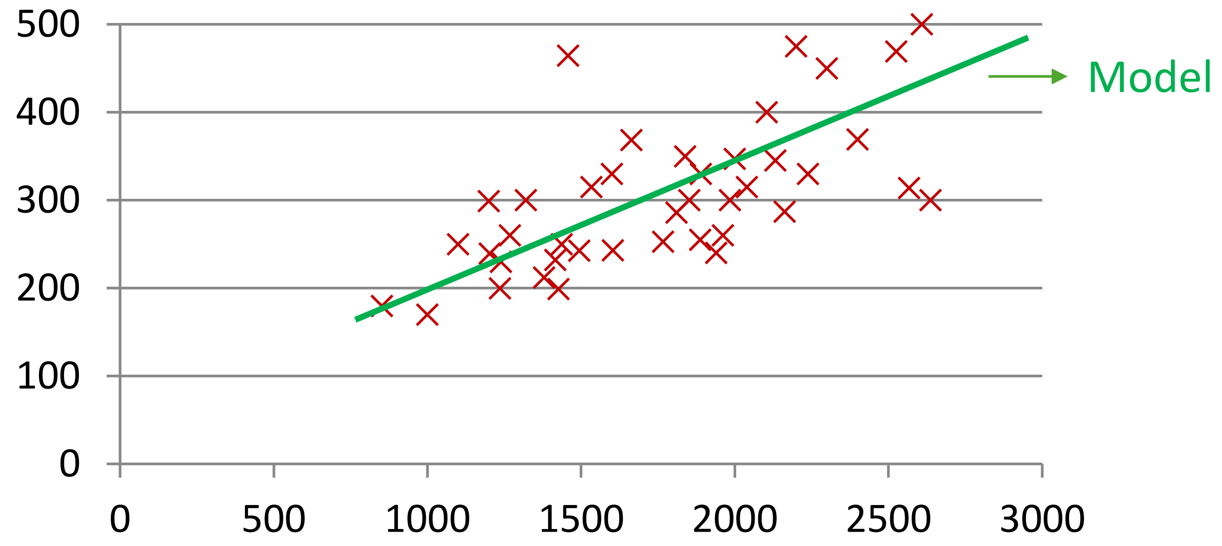


BMI x_0	Daily Exercise x_1	Class y
22.5	2.5	Healthy
23.0	1.5	Healthy
30	1	Unhealthy
⋮	⋮	⋮

Simple Models for Classification and Regression

- ML is the paradigm of approximating a function (model) from data.

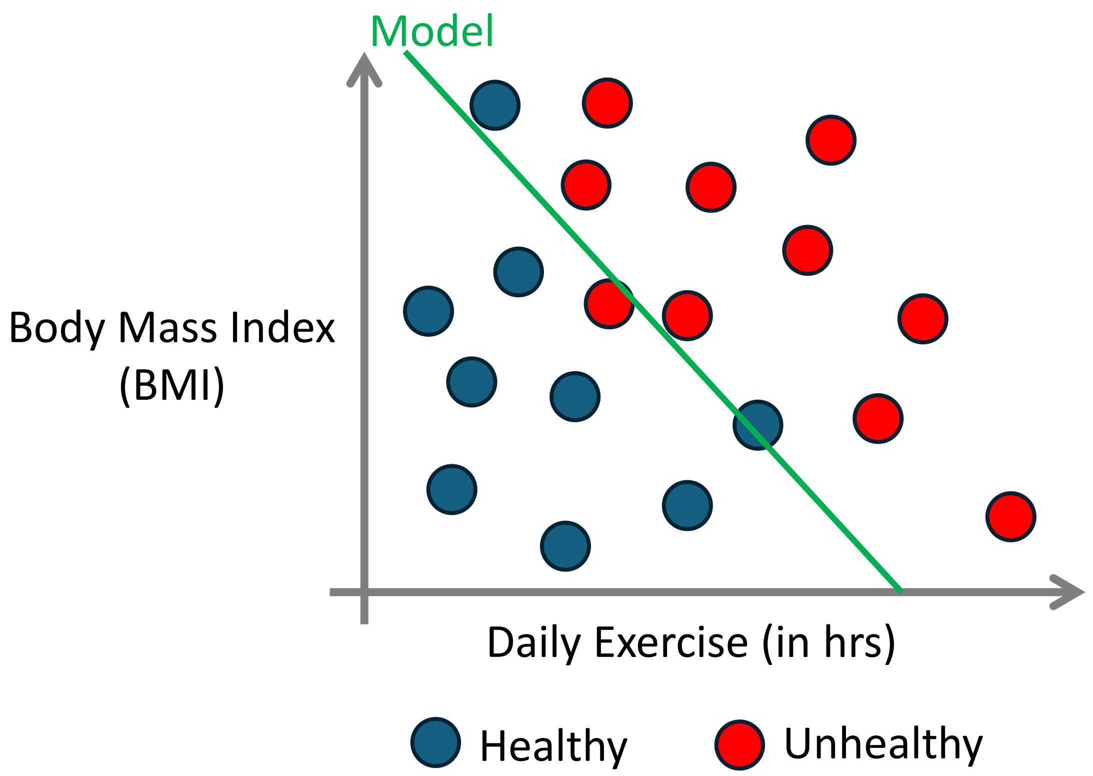
Regression



Regression: Predict exact value

$$x = 1000$$
$$y = ?$$

Classification

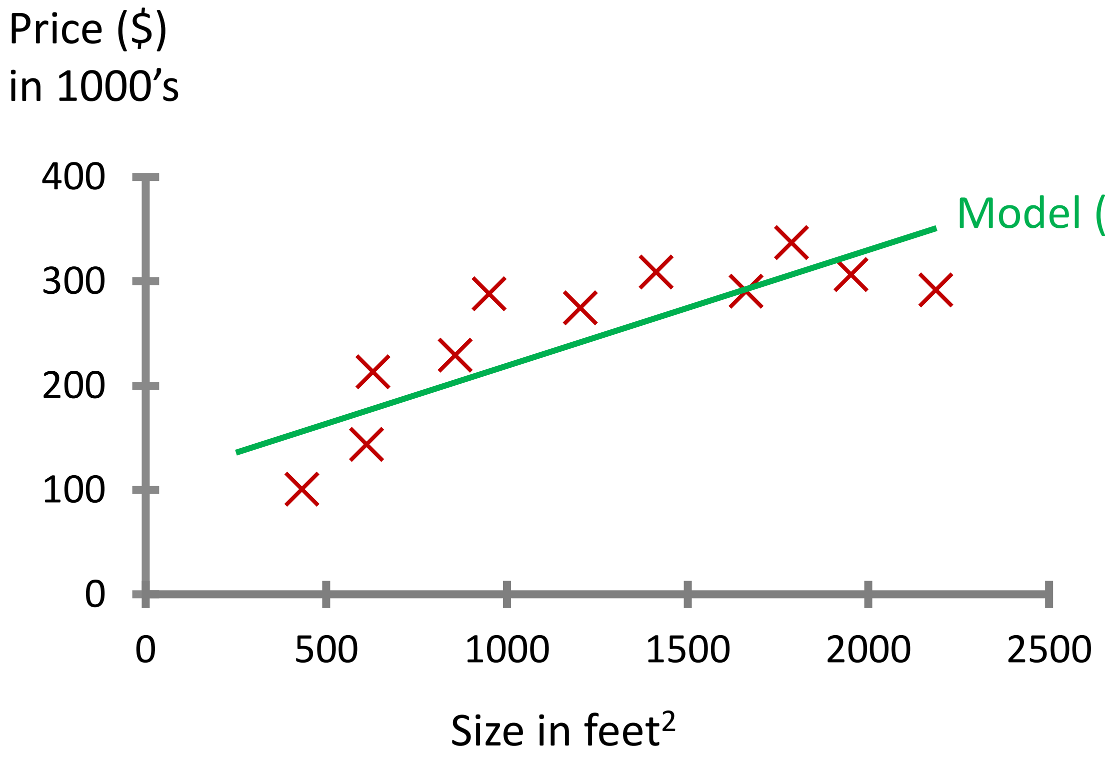


Classification: Predict class label

BMI = 18.5, Daily exercise = 1

Healthy OR **Unhealthy**

Parameters of the model



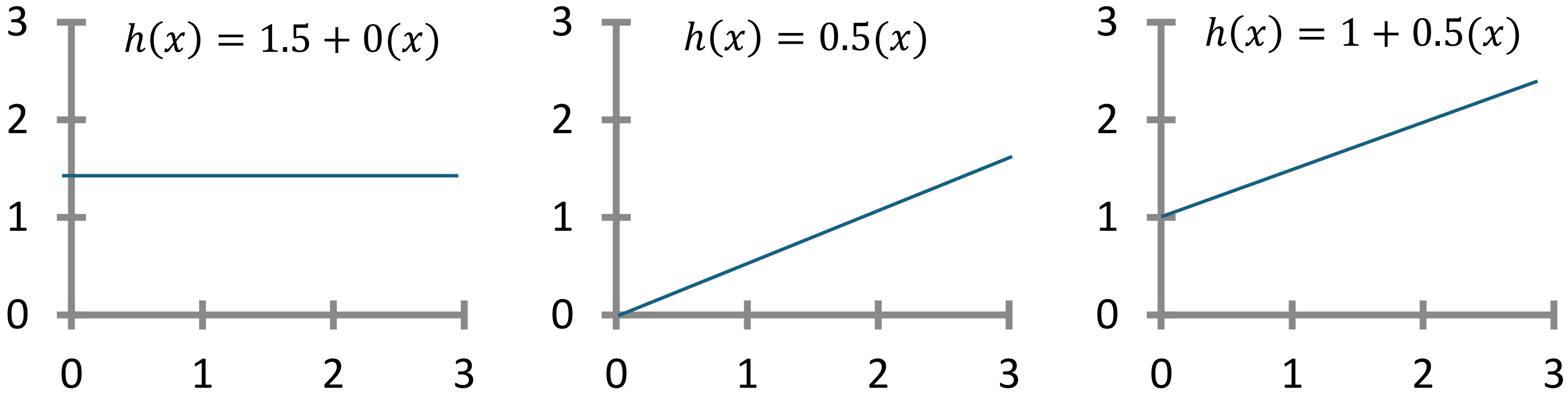
Model (Hypothesis) : $h_\theta(x) = \theta_0 + \theta_1 x$

Parameters (θ_i)
$i = 0$; bias
$i \neq 0$; weight

How to choose $\theta_i s$?

Learning Parameters

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



$\theta_0 = 1.5$, $\theta_1 = 0$

$\theta_0 = 0$, $\theta_1 = 0.5$

$\theta_0 = 1$, $\theta_1 = 0.5$

Train, Validation and Test Sets

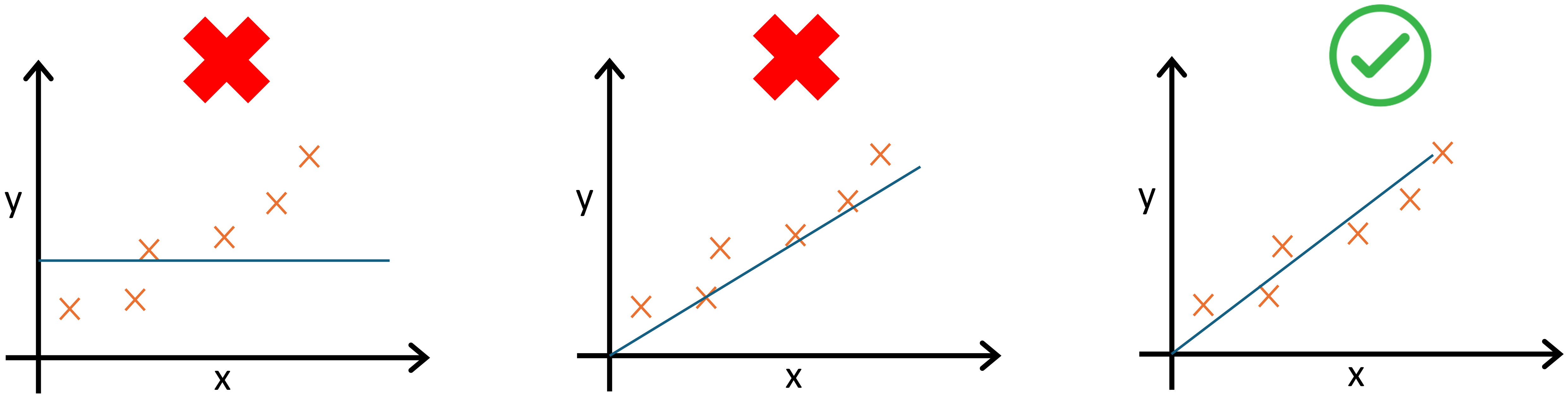
Size in feet² (x)	**Price ($) in 1000's (y)**
2104	460
1416	232
1534	315
⋮	⋮

Training Set (D_{train})
data used for finding the optimal parameters

Validation Set (D_{val})
data used to diagnose the training stage

Test Set (D_{test})
data used for testing the model's ability to correctly predict on unseen data

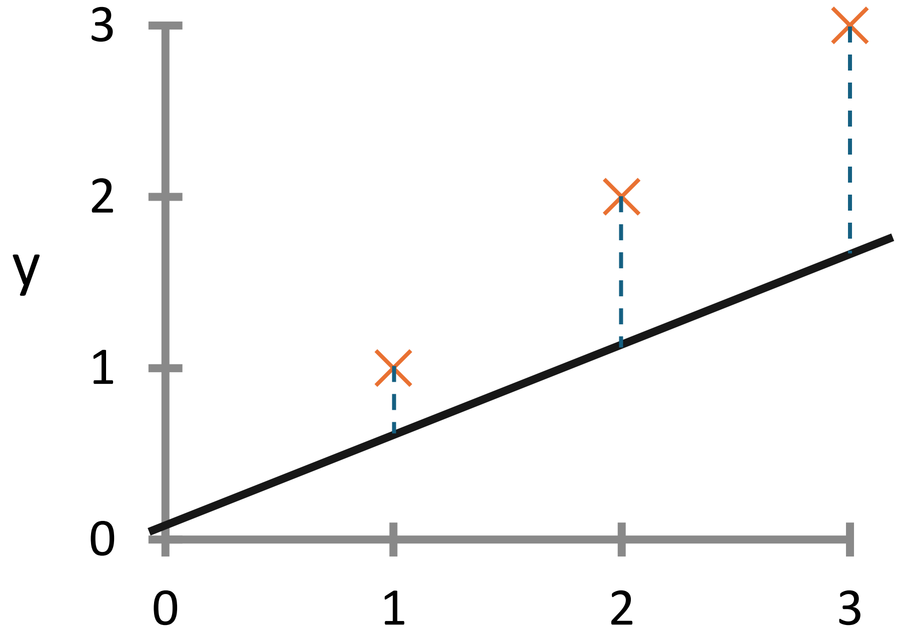
Learning Parameters



Choose parameters ($\boldsymbol{\theta}$'s) so that $h_{\boldsymbol{\theta}}(x)$ ($\hat{y}$) is close to $\boldsymbol{y}$ for our training examples (D_{train})

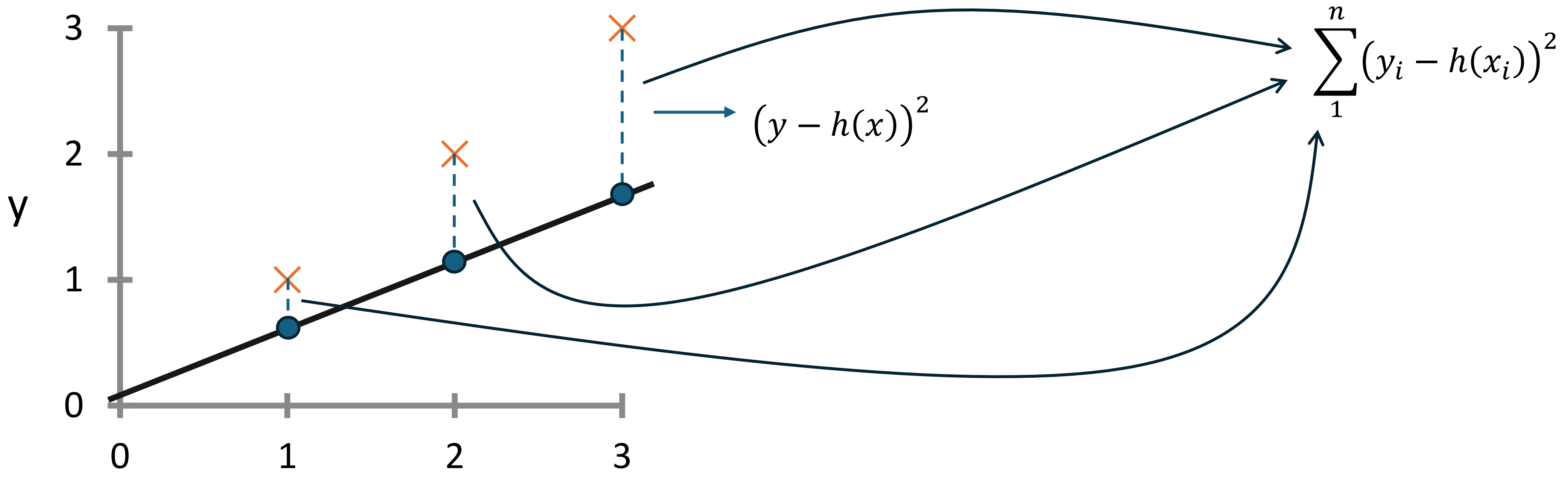
Learning Parameters

- Choose parameters (θ_i) so that $h(x)$ is close to y for our training examples
- A loss function $L\big(y, h_\theta(x)\big)$ quantifies the gap between the actual data and the model.



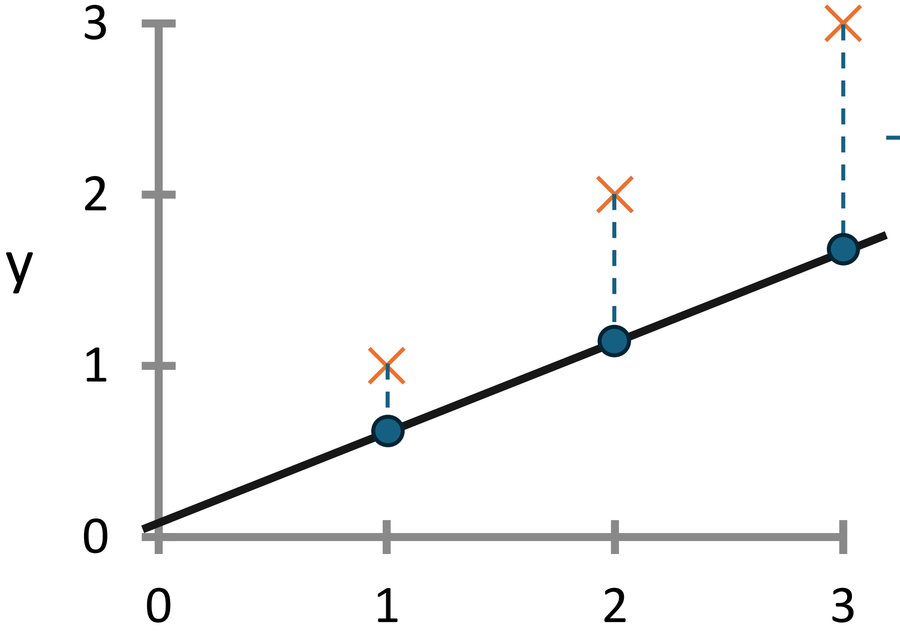
Learning Parameters

- A loss function $L\big(y, h_\theta(x)\big)$ quantifies the gap between the actual data and the model.



Learning Parameters

- A loss function $L\big(y, h_\theta(x)\big)$ quantifies the gap between the actual data and the model.



$\big(y - h(x)\big)^2$

$$\sum_1^n \big(y_i - h(x_i)\big)^2$$

Average out the sum

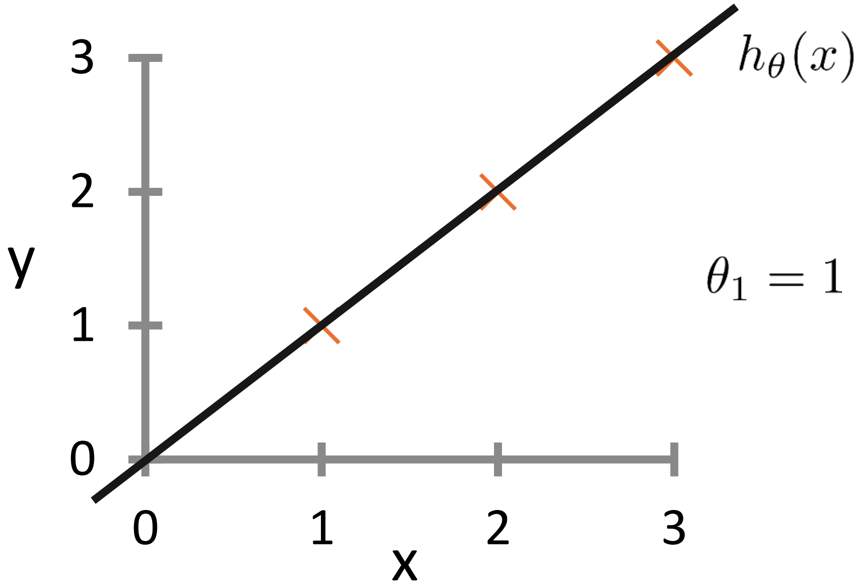
$$L\big(y, h(x)\big) = \frac{1}{n}\sum_1^n \big(y_i - h(x_i)\big)^2$$

Goal:
Minimize the loss

Minimizing loss = finding optimal θs

Learning Parameters

- Example:
  - Hypothesis: $h_\theta(x) = \theta_1 x$
  - Parameters: θ_1
  - Loss: $L\big(y, h(x)\big) = \frac{1}{n}\sum_1^n\big(y_i - h(x_i)\big)^2$



- $x = 1;\ y = 1, h(1) = 1$

$$y_1 - h(x_1) = 1 - 1 = 0$$

- $x = 2;\ y = 2, h(2) = 2$

$$y_2 - h(x_2) = 2 - 2 = 0$$

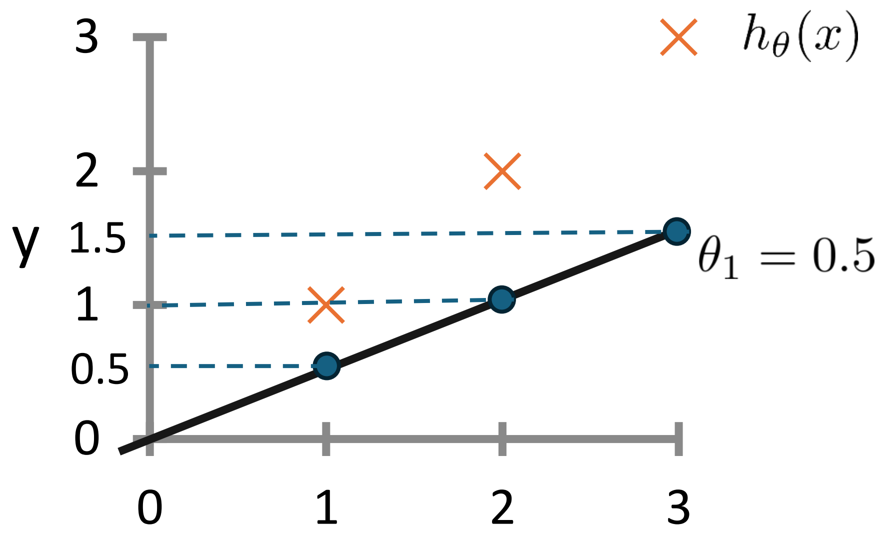
- $x = 3;\ y = 3, h(3) = 3$

$$y_3 - h(x_3) = 3 - 3 = 0$$

$$L\big(y, h_{\theta=1}(x)\big) = \frac{1}{3}(0^2 + 0^2 + 0^2) = 0$$

Learning Parameters

- Example:
  - Hypothesis: $h_\theta(x) = \theta_1 x$
  - Parameters: θ_1
  - Loss: $L\big(y, h(x)\big) = \frac{1}{n}\sum_1^n \big(y_i - h(x_i)\big)^2$



  - $x = 1;\ y = 1, h(1) = 0.5$

$$y_1 - h(x_1) = 1 - 0.5 = 0.5$$

  - $x = 2;\ y = 2, h(2) = 1$

$$y_2 - h(x_2) = 2 - 1 = 1$$

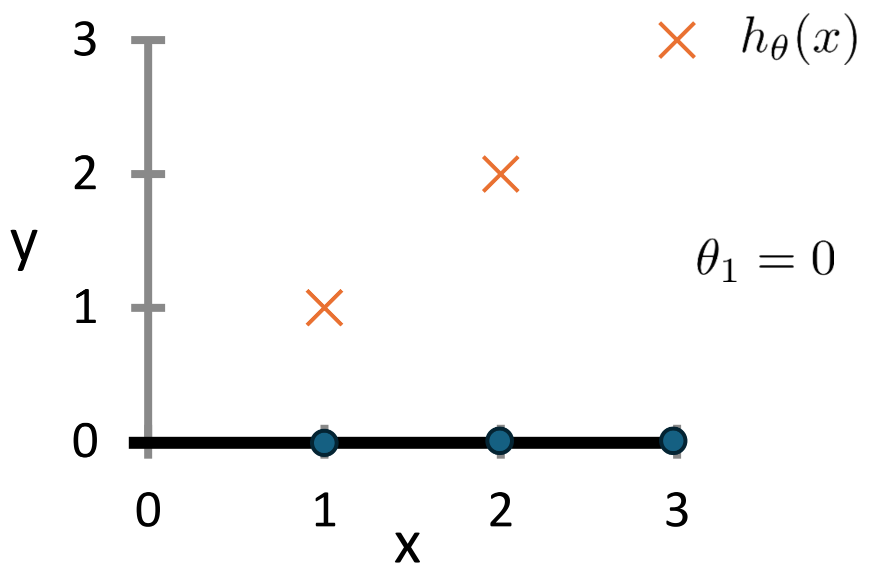
  - $x = 3;\ y = 3, h(3) = 1.5$

$$y_3 - h(x_3) = 3 - 1.5 = 1.5$$

$$L\big(y, h_{\theta=0.5}(x)\big) = \frac{1}{3}(0.5^2 + 1^2 + 1.5^2) = \frac{7}{6} = 1.12$$

Learning Parameters

- Example:
  - Hypothesis: $h_\theta(x) = \theta_1 x$
  - Parameters: θ_1
  - Loss: $L\big(y, h(x)\big) = \frac{1}{n}\sum_1^n \big(y_i - h(x_i)\big)^2$



  - $x = 1;\ y = 1, h(1) = 0$

$$y_1 - h(x_1) = 1 - 0 = 1$$

  - $x = 2;\ y = 2, h(2) = 0$

$$y_2 - h(x_2) = 2 - 0 = 2$$

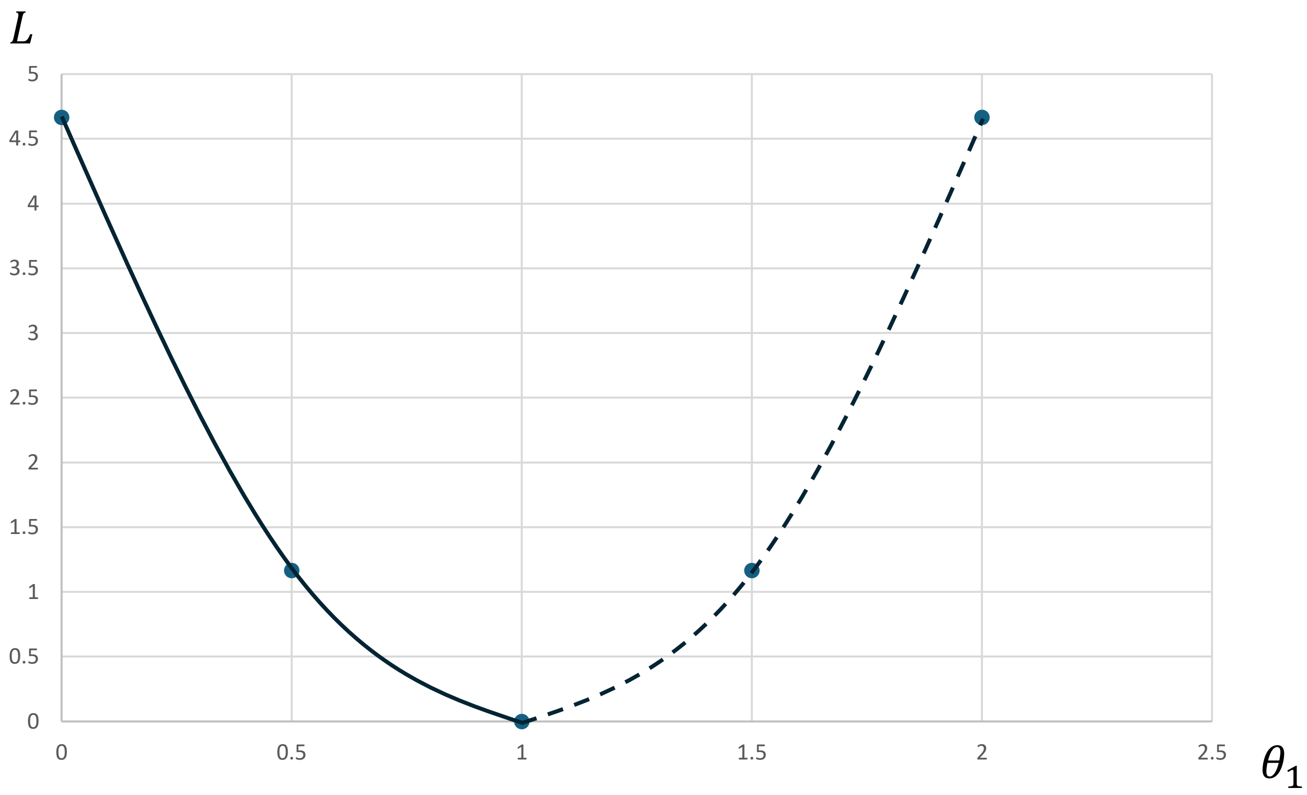
  - $x = 3;\ y = 3, h(3) = 0$

$$y_3 - h(x_3) = 3 - 0 = 3$$

$$L\big(y, h_{\theta=0}(x)\big) = \frac{1}{3}(1^2 + 2^2 + 3^2) = 4.67$$

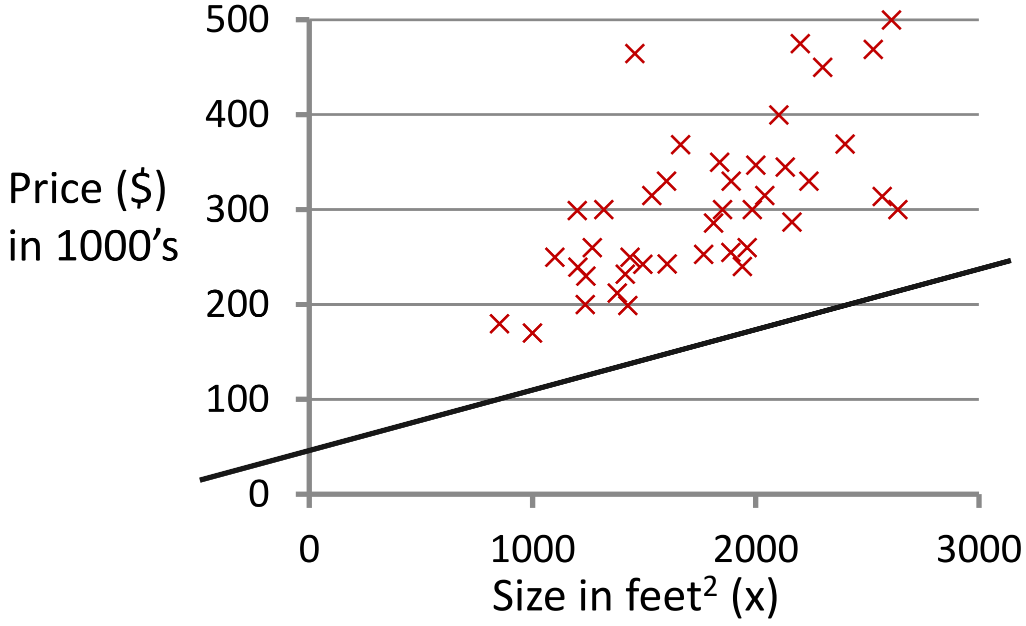
Learning Parameters

- Example:
  - $L(\theta = 1) = 0$
  - $L(\theta = 0.5) = 1.12$
  - $L(\theta = 0) = 4.67$



Learning Parameters

- Example:



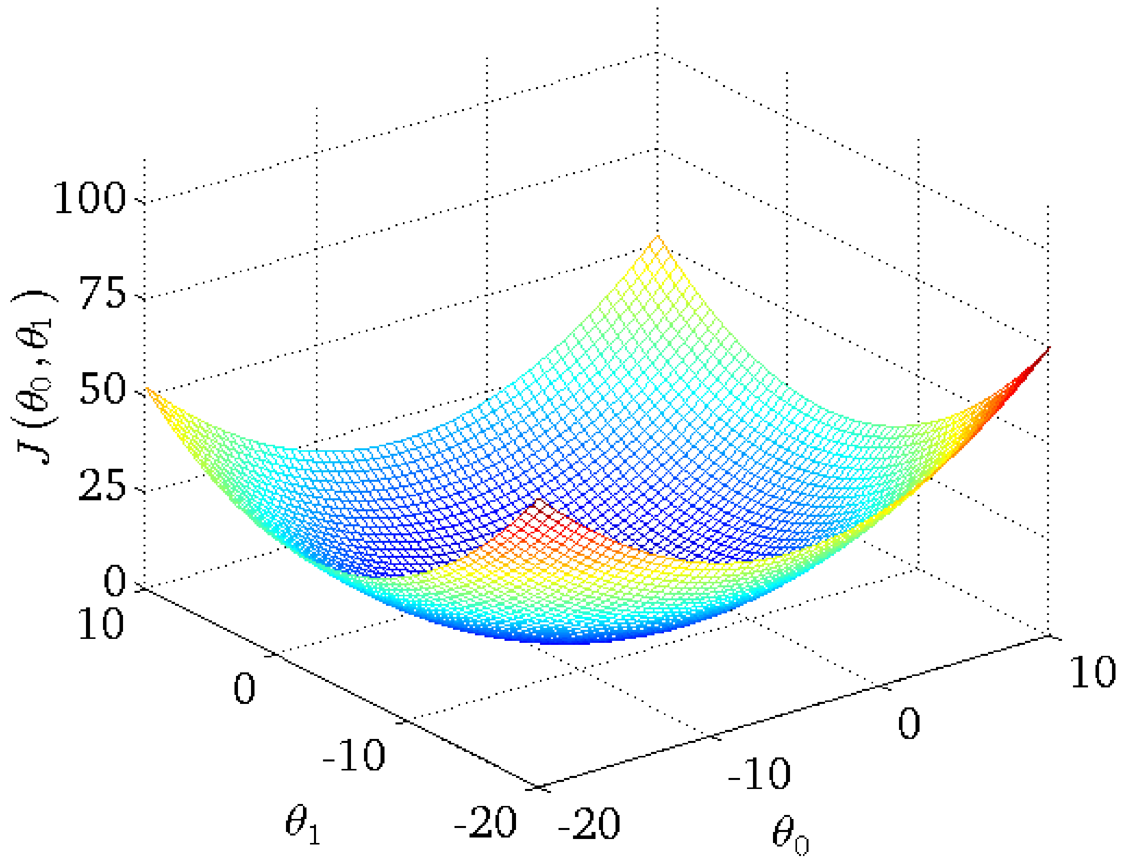
$$h_\theta(x) = 50 + 0.06x$$

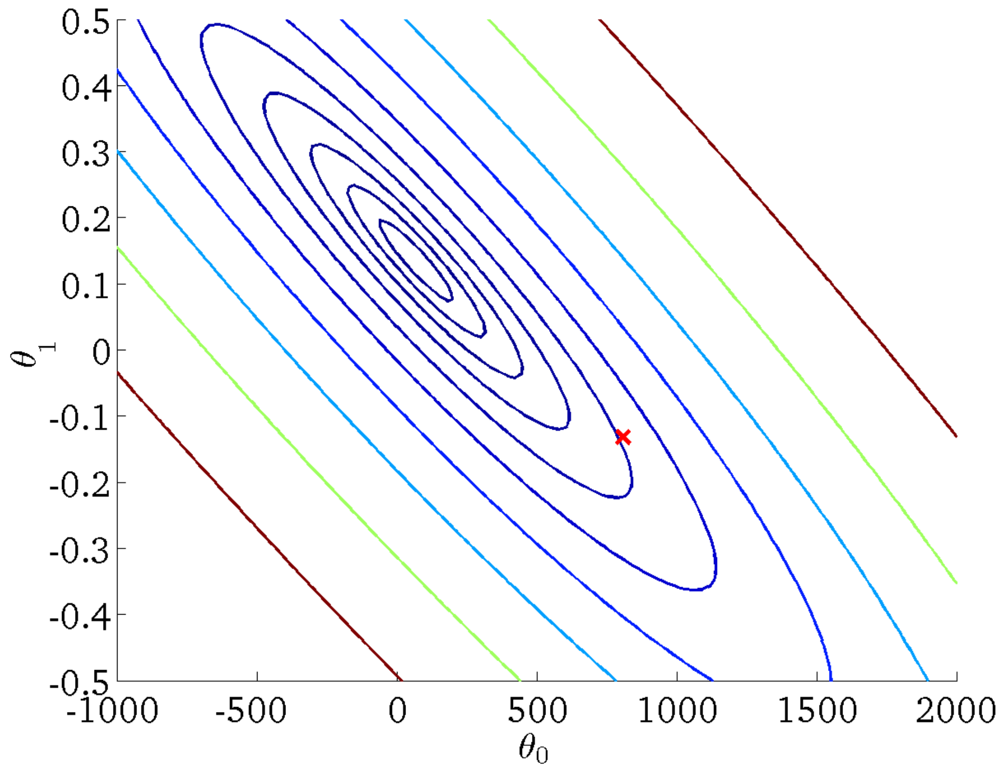
$$L\left(y, h_{\theta=(50,0.06)}(x)\right) = J(\theta_0, \theta_1)$$

Learning Parameters

- Example:

$$L\left(y, h_{\theta=(800,-0.13)}(x)\right)$$

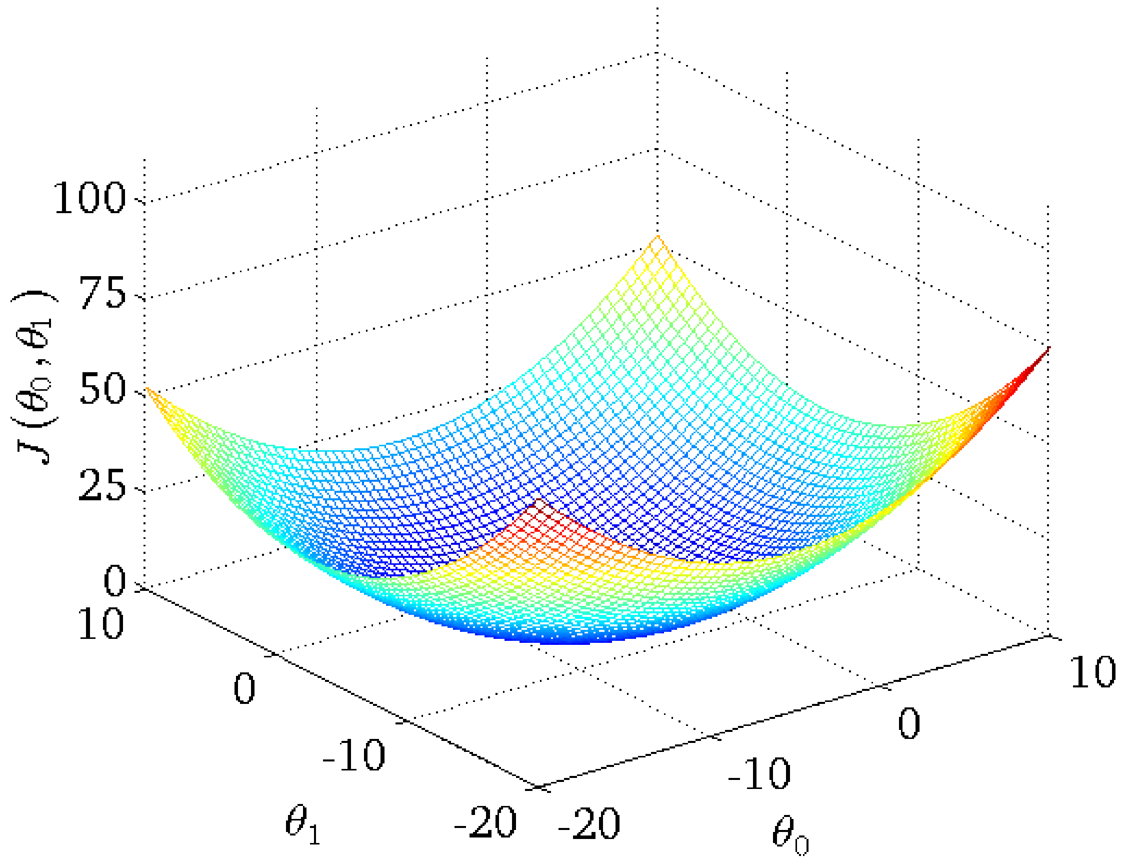


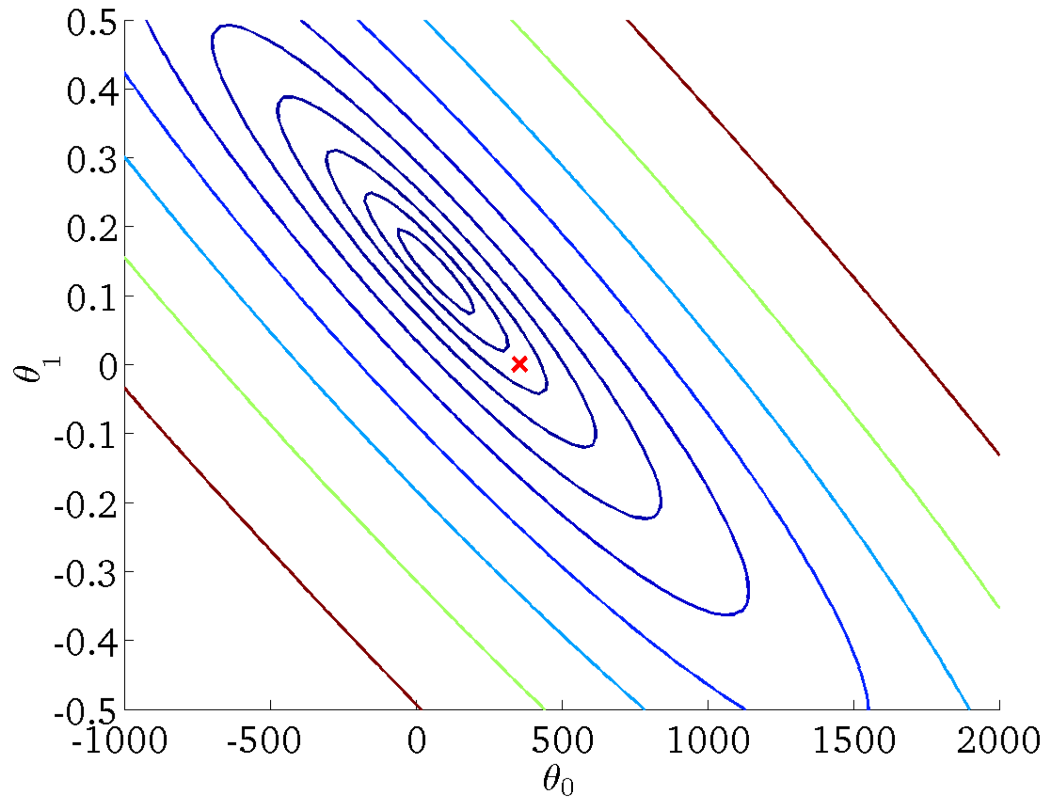


Learning Parameters

- Example:

$$L\left(y, h_{\theta=(360,0)}(x)\right)$$

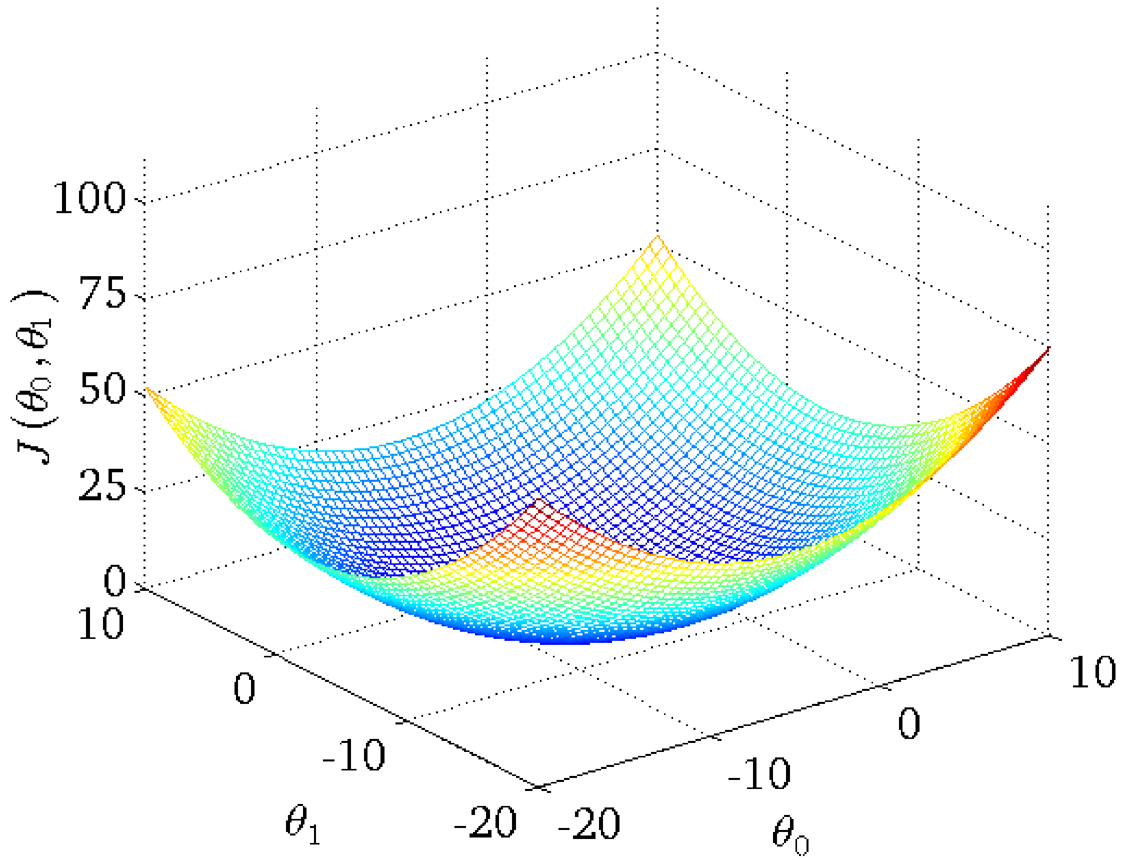


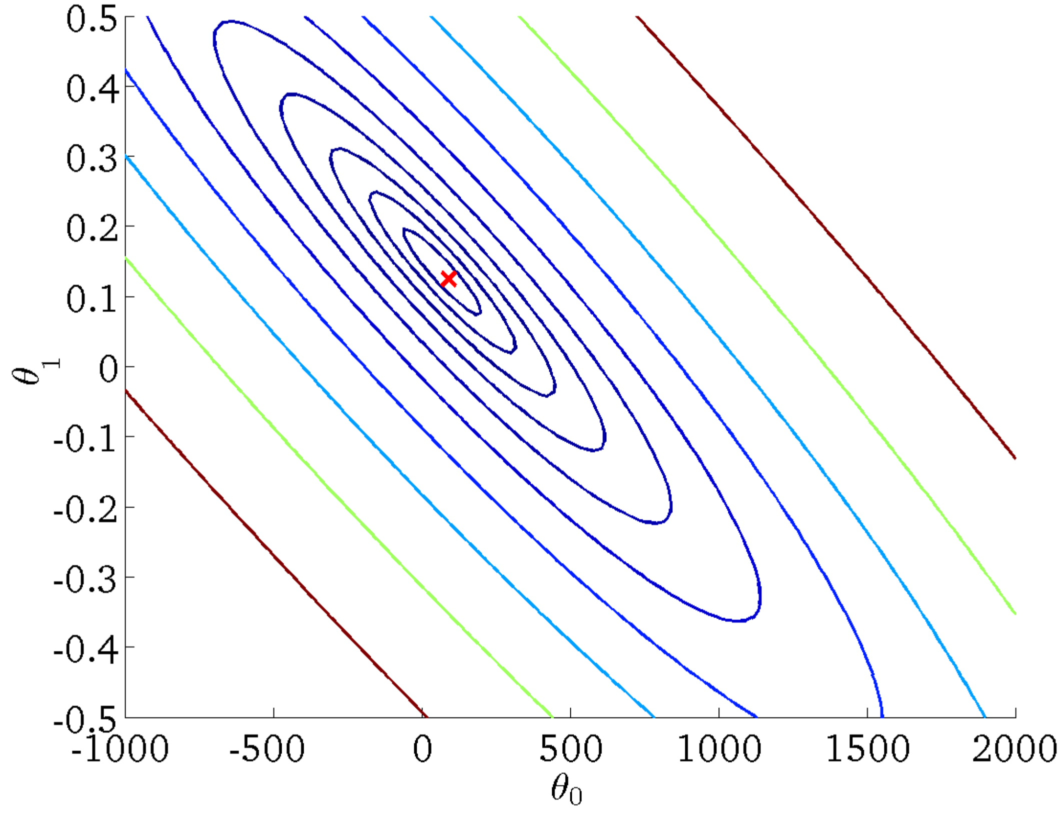


Learning Parameters

- Example:

$$L\left(y, h_{\theta=(250,0.13)}(x)\right)$$



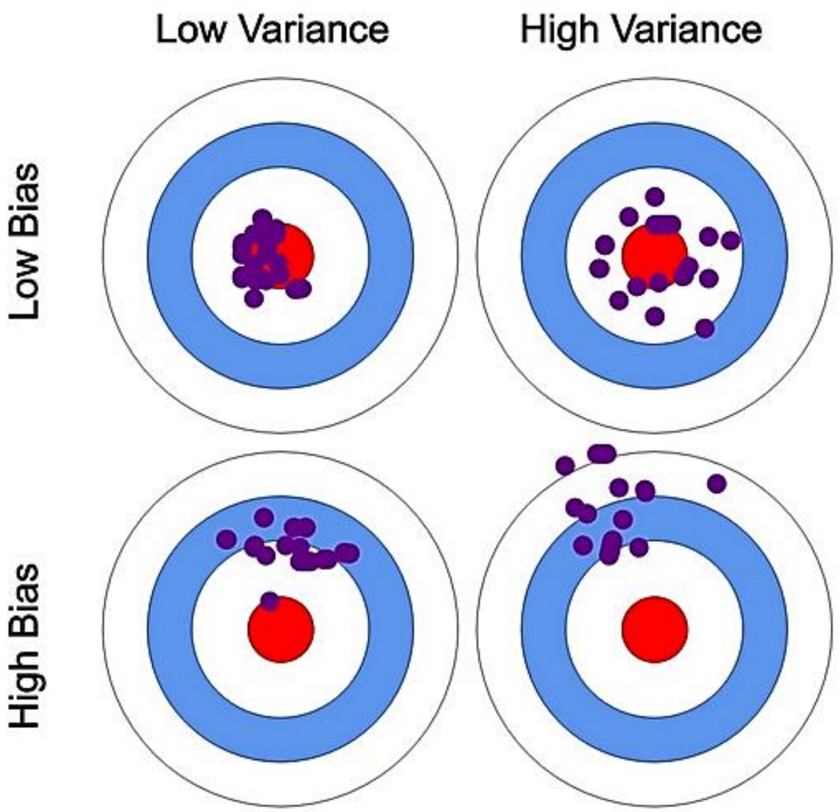


Generalization

- How well a model generalizes can be characterized by the difference between its performance on data it has seen vs not seen
- If a model is made more “complex”, it might be able to learn more “complex patterns” but we also risk simply memorizing the training data instead of truly learning anything from it.

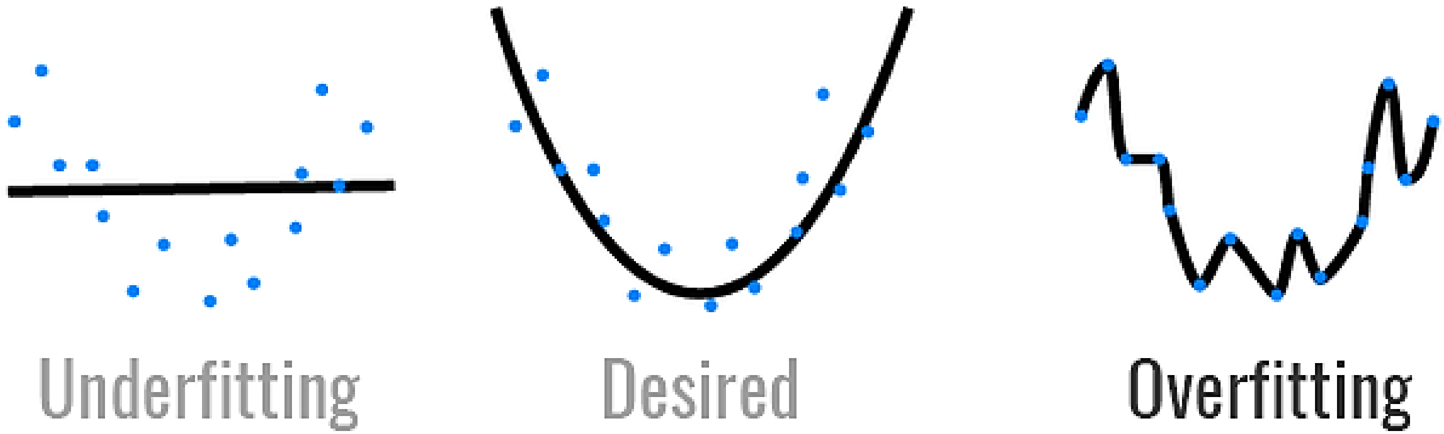
Generalization

- Bias & Variance
  - **Bias:**
    - Error introduced due to oversimplified models (e.g., linear models on non-linear data). It causes systematic errors in predictions.
  - **Variance:**
    - How sensitive is the model to changes in the training data
    - Small changes in dataset → large changes in the model and its predictions
  - A good models needs to be both firm and flexible: able to capture varying and complex data yet robust enough to generalize beyond just the training samples.



Generalization

- Overfitting & Underfitting



- Variance is too high ⇒ Overfitting
  - too little data
  - too complex model (function) class
- Bias is too high ⇒ Underfitting
  - Insufficiently complex model (function) class

References

- Alammar, J., & Grootendorst, M. Hands-On Large Language Models: Language Understanding and Generation. O'Reilly Media.
- UC Berkeley. Modern Computer Vision: Introduction to Machine Learning. Course lecture slides.
- Ng, A. Machine Learning. Coursera.