Applied Artificial Intelligence

Session 6: Probability and Statistics for AI and Machine Learning

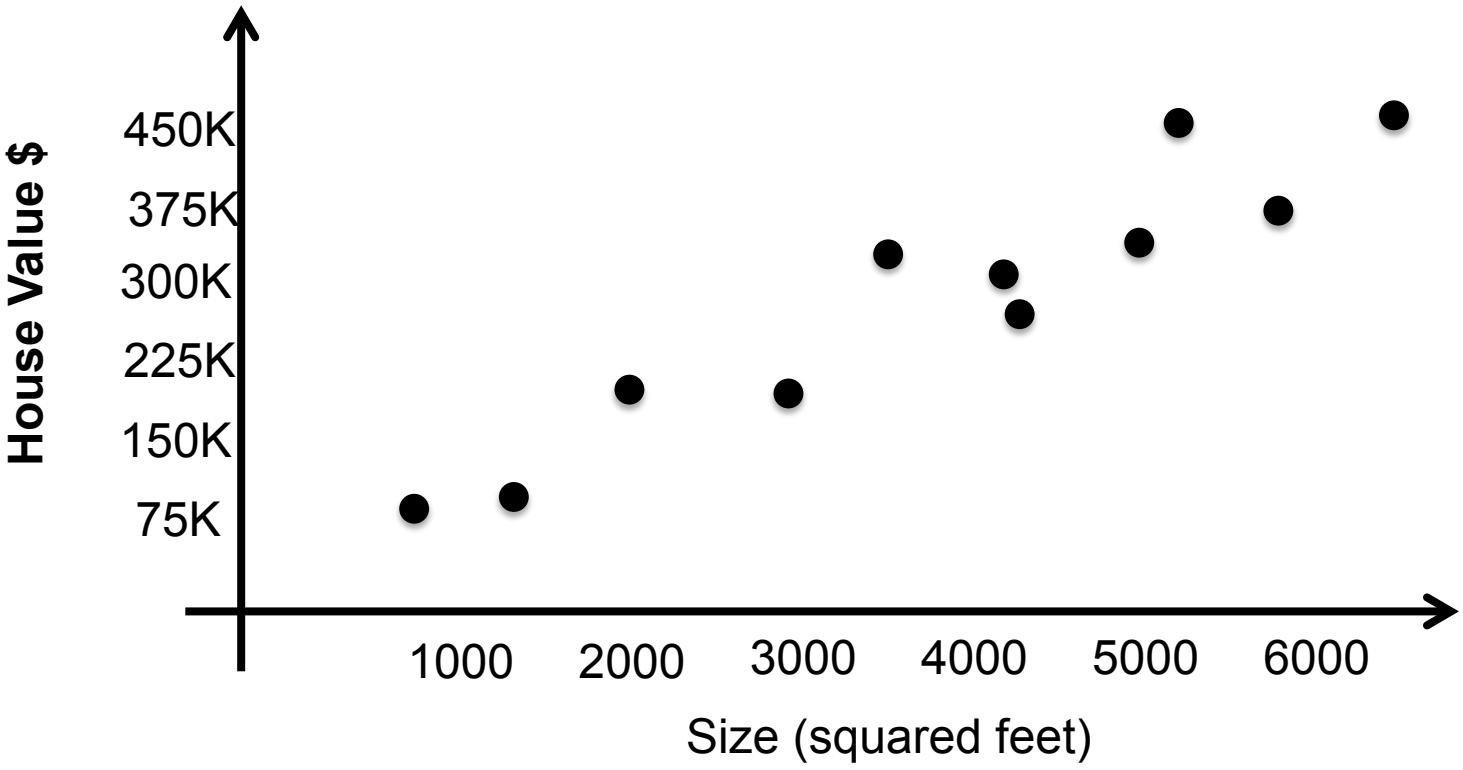
Fall 2018

NC State University

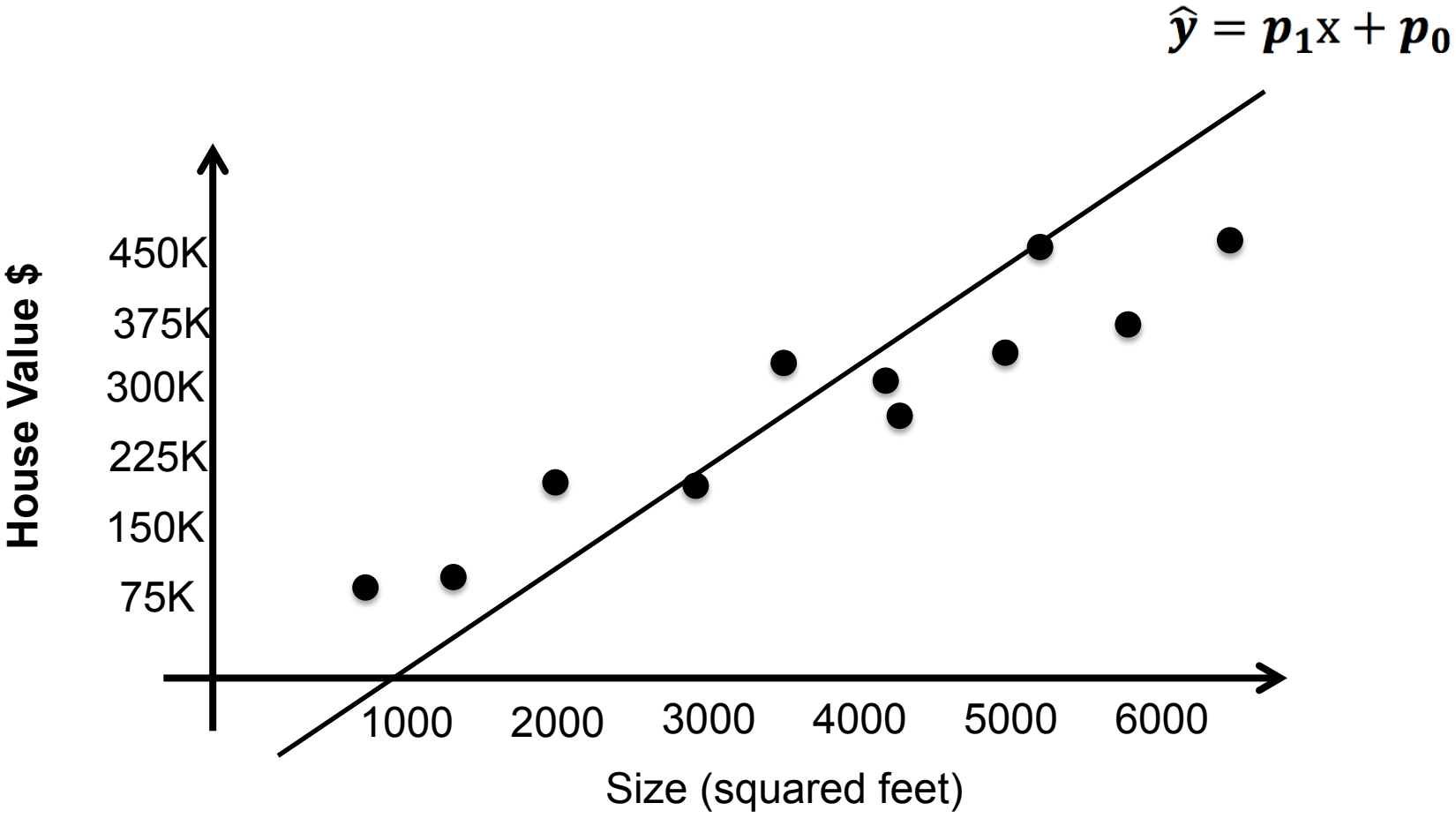
Instructor: Dr. Behnam Kia

Course Website: https://appliedai.wordpress.ncsu.edu/

Example: House Values

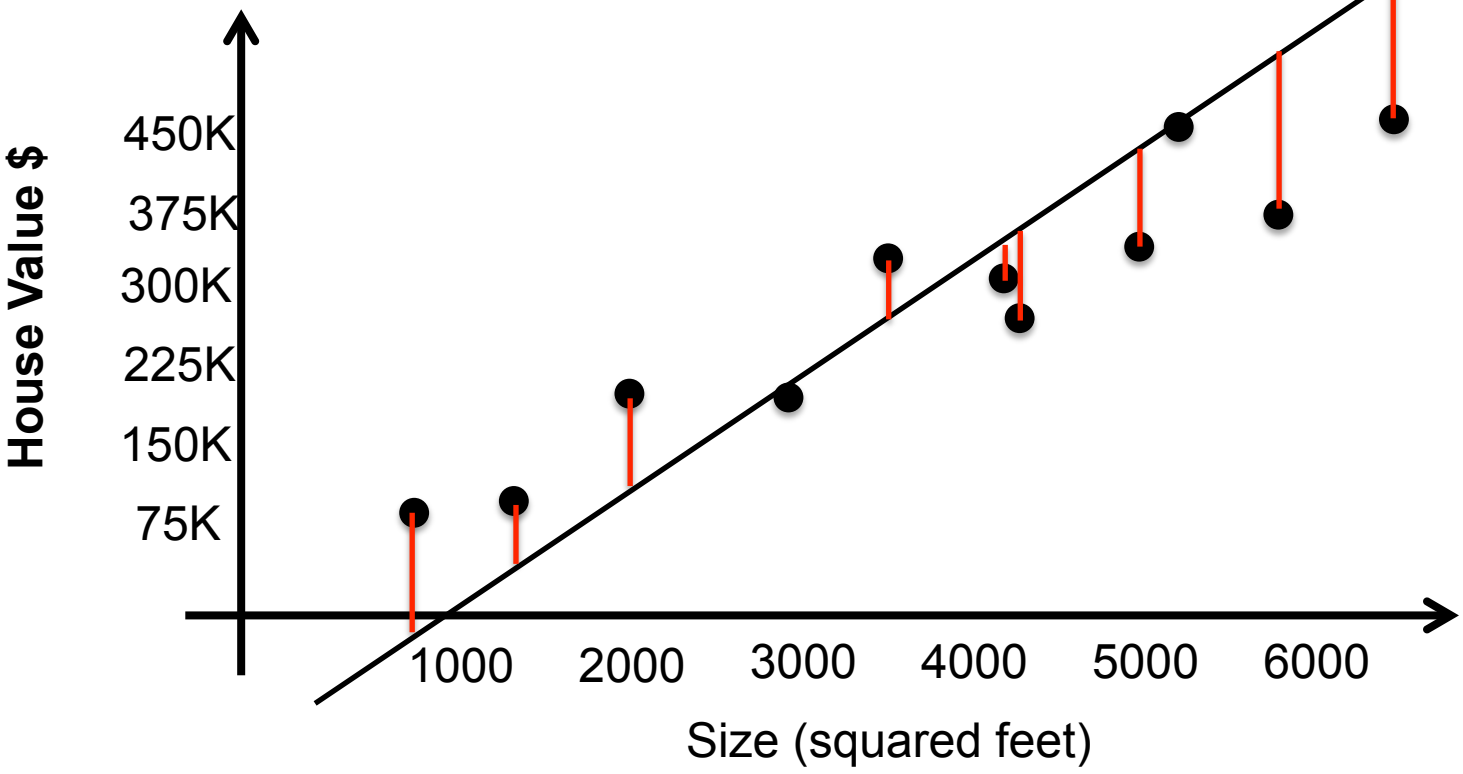


Example: House Values



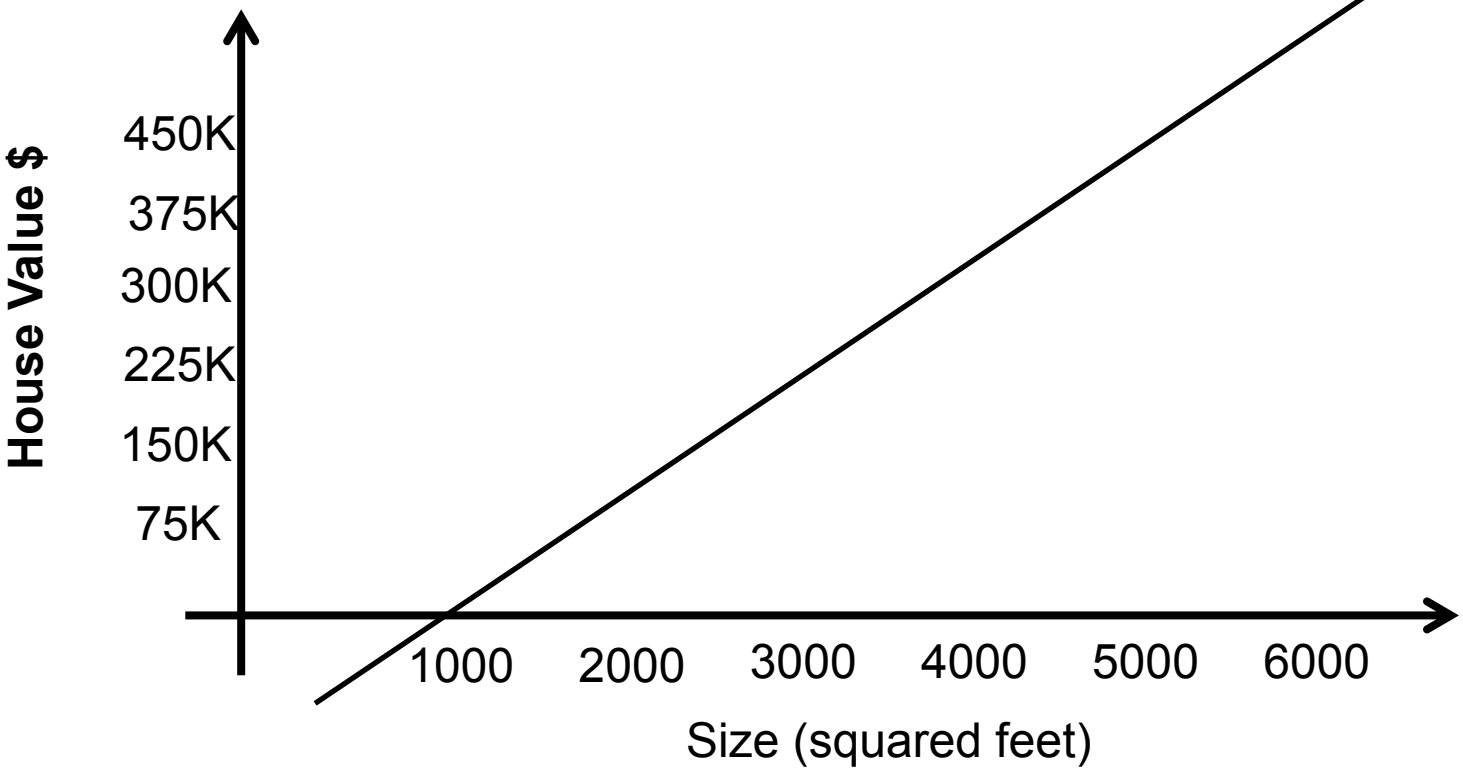
Example: House Values

$$\hat{y} = p_1 x + p_0$$

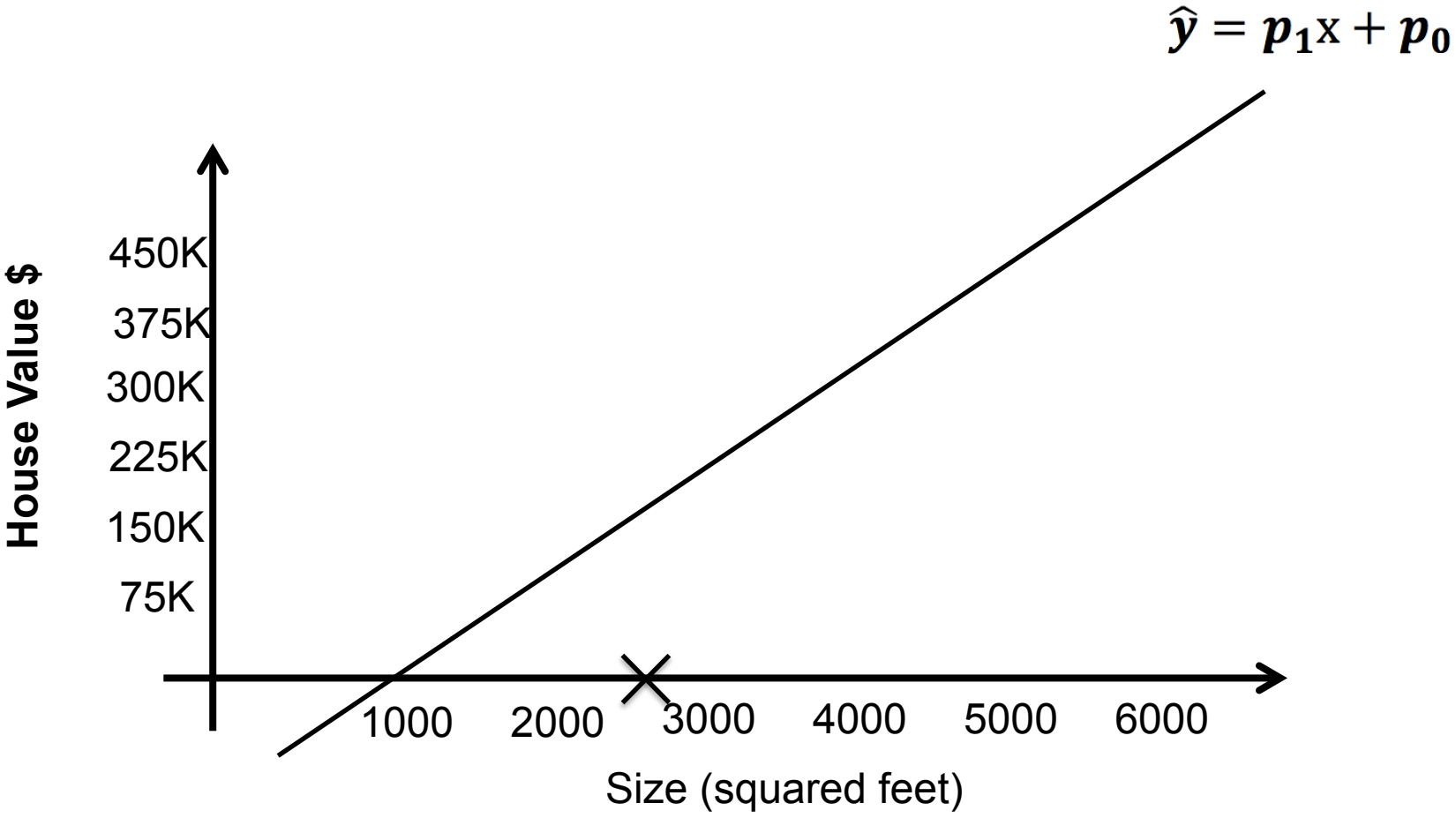


Example: House Values

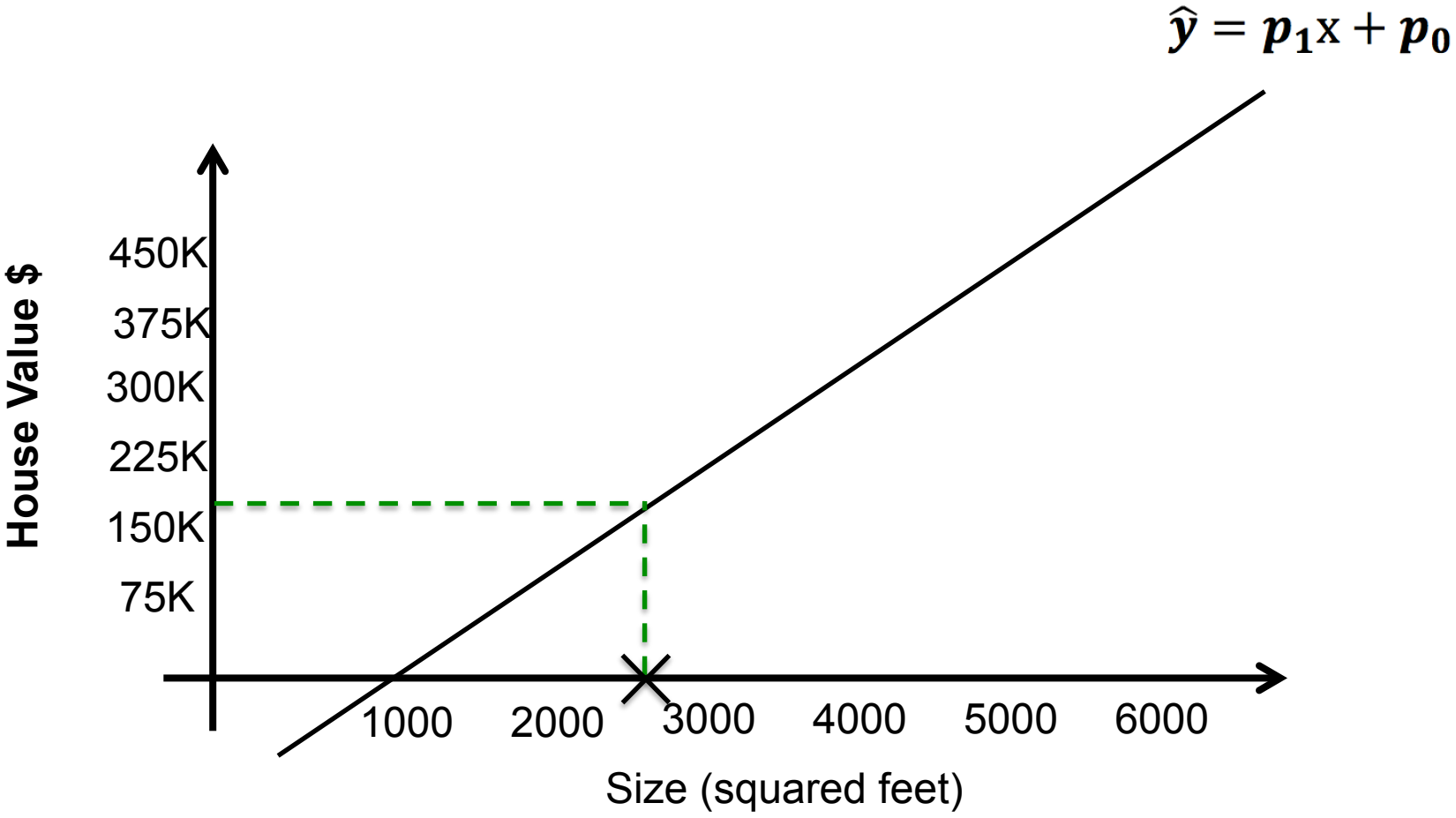
$$\hat{y} = p_1 x + p_0$$



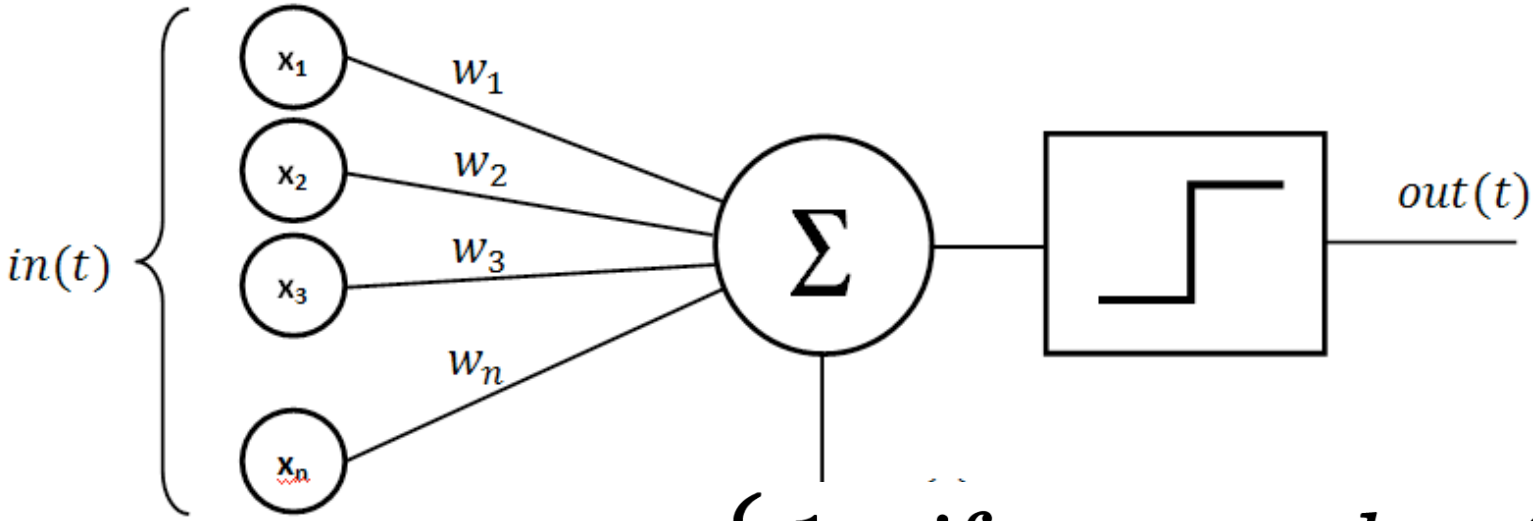
Example: House Values



Example: House Values



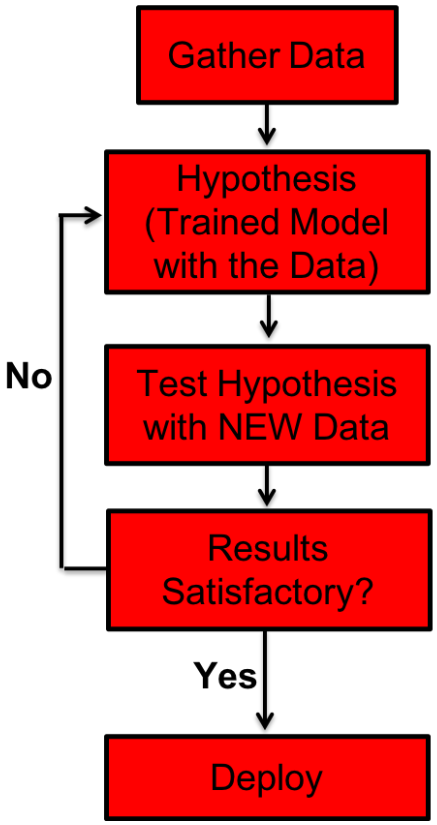
Perceptron: A Computational Neuron Model

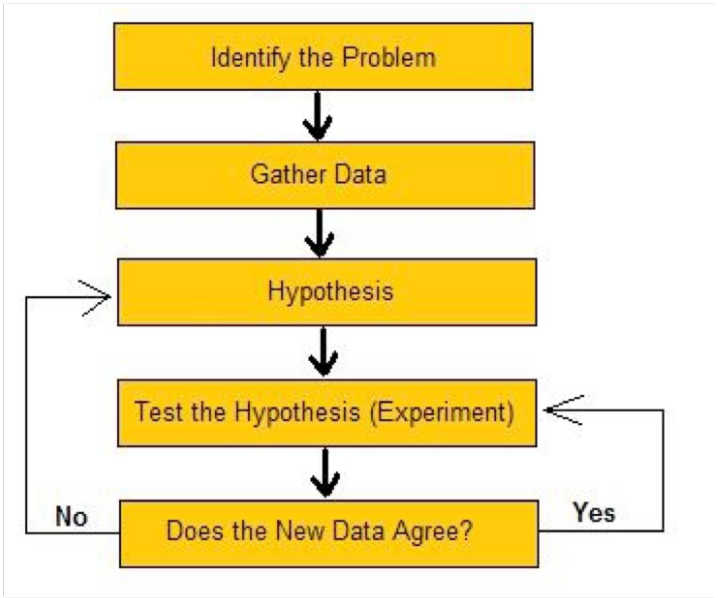


$$f(x) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

Questions and Discussions About the GA + Perceptron Project

- **Did we test our hypothesis?**

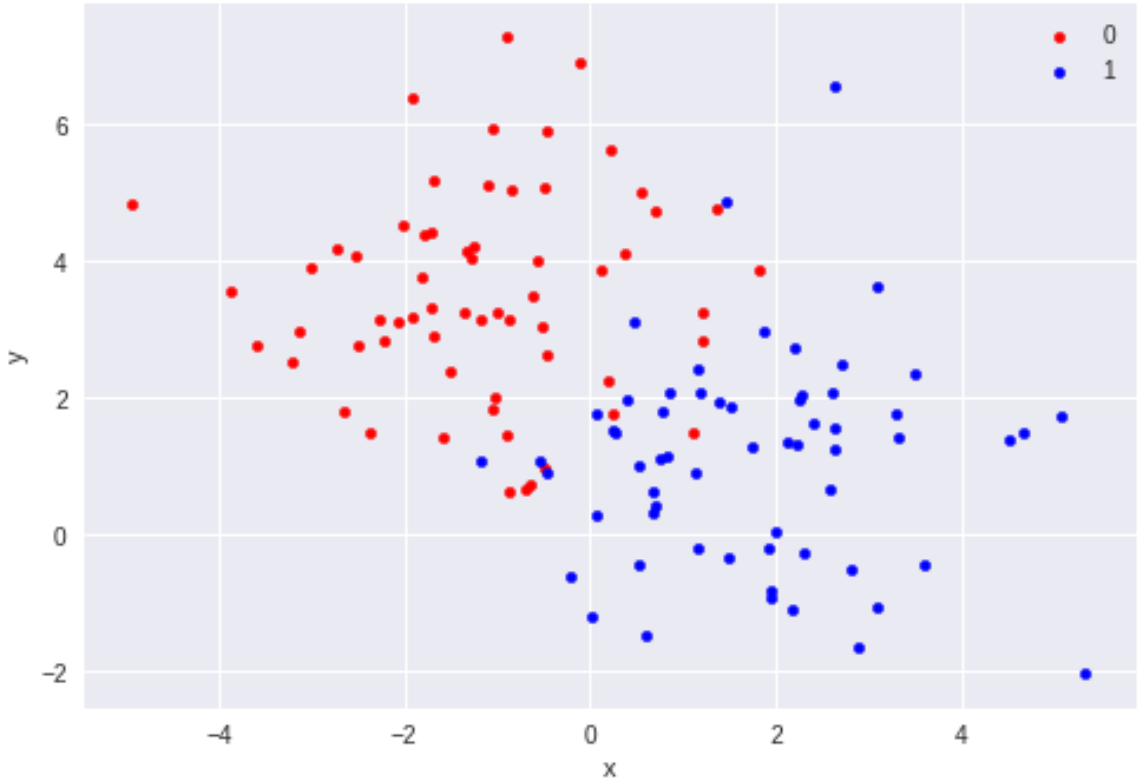




Scientific Method

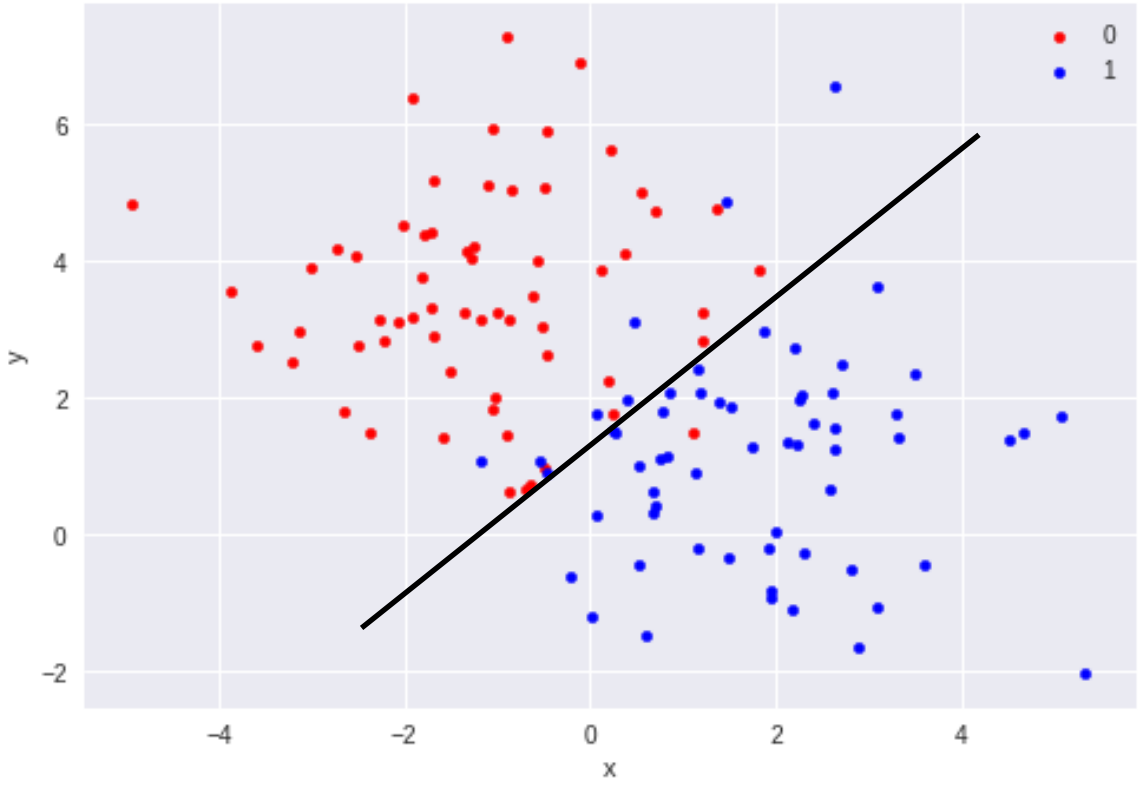
Questions and Discussions About the GA + Perceptron Project

- Did we test our hypothesis?



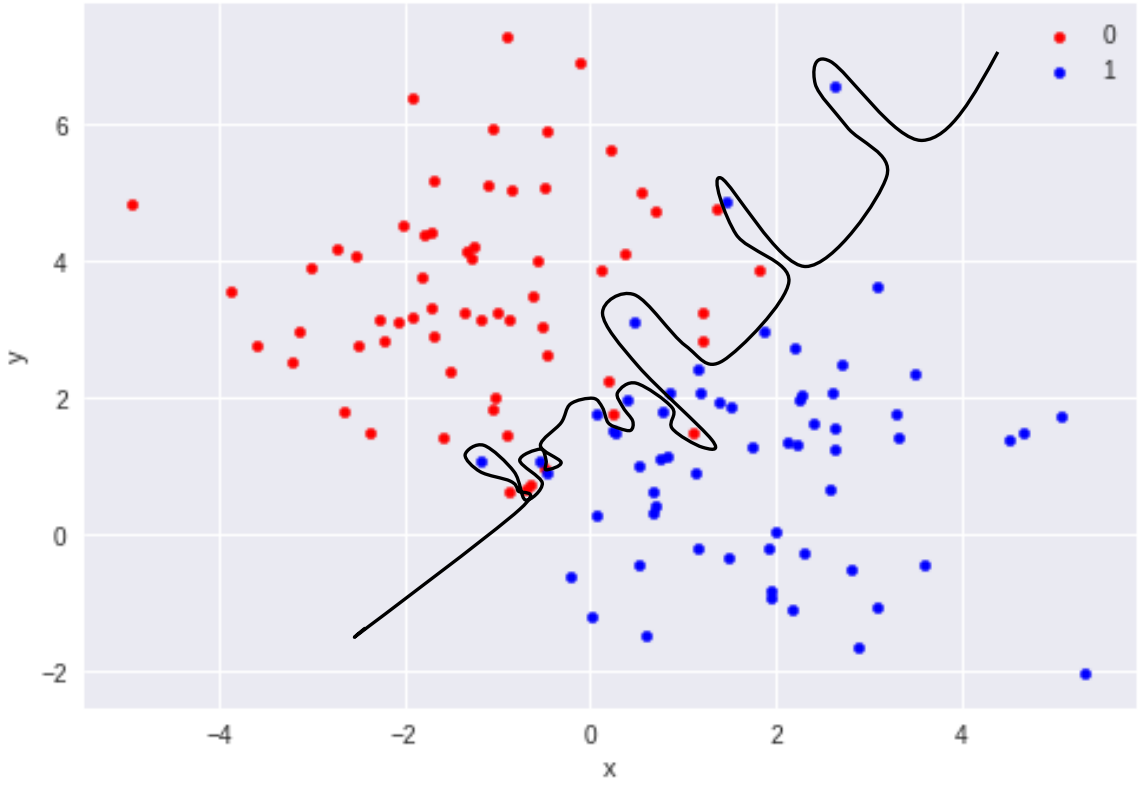
Questions and Discussions About the GA + Perceptron Project

- Did we test our hypothesis?



Questions and Discussions About the GA + Perceptron Project

- Did we test our hypothesis?



Probability and Statistics for AI and Machine Learning

- What is Probability and Statistic?
- Why do we need probability and statistics for AI and Machine Learning?
- What are the main results of Probability and Statistics that we need in AI and Machine Learning?

What is Probability

- Probability theory is a mathematical framework for representing and working with uncertain statements.
  - It quantifies uncertainty (degree of belief):

    P("It is a **Cat**")=0.75
  - And it provides axioms and rules to reason and derive new uncertain statements.
- P("It is a **Cat**" | "It's 10lbs")=P("It's10lbs" | "It is a **Cat**") *P("It is a **Cat**")/P("It's 10lbs")

What is Statistics?

- Statistics deals with data. Generally speaking, the goal of statistics is to draw conclusions (making inferences) from data.

Why Do We Need Probability and Statics for AI and Machine Learning?

- In AI and machine learning we deal with uncertain quantities:
  - Incomplete observability: limited data and view, noisy data, etc.
  - Incomplete models (noisy models): no perfect learning models. We make assumptions for each model, and it comes with drawbacks.
  - Inherent stochasticity of the system we model: Quantum Mechanics, Chaos Theory, flipping a coin.

Reading Assignment

- Please refresh your knowledge of probability and statistics.
- Please see reading assignment posted on the website for complete instructions.

Conditional Probability

- What is conditional probability?

P($y|x$)?

Conditional Probability

- What is conditional probability?

$P(y|x)=P(y,x)/P(x)$

Bayes Rule

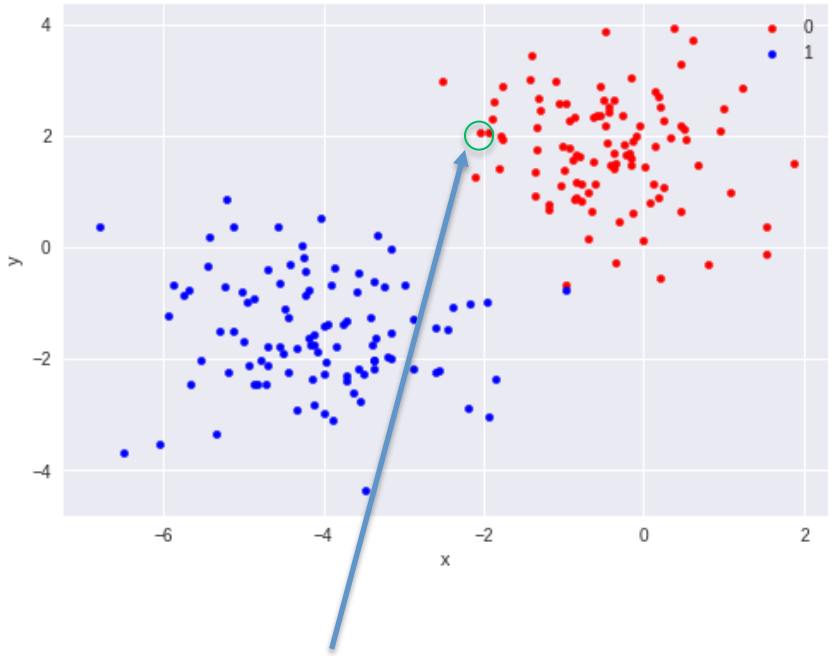
- $P(y|x)=P(x|y)P(y)/P(x)$?

Why conditional probability matters for AI and machine learning?

Why conditional probability matters for AI and machine learning?

$$f^{BO}(x) = \underset{y}{\operatorname{argmax}}\ p(y \mid x)$$

Why conditional probability matters for AI and machine learning?



$$f^{BO}(x) = \underset{y}{\operatorname{argmax}}\ p(y \mid x)$$

$$f^{BO}(x) = \begin{cases} 0 & if\ \ p(0 \mid (-2,2)) > p(1 \mid (-2,2)) \\ 1 & if\ \ p(1 \mid (-2,2)) \geq p(0 \mid (-2,2)) \end{cases}$$

Bayes Optimal Classifier

- The Bayes Optimal Classifier $f^{(BO)}$ achieves minimal zero/one error of any deterministic classifier.

$$f^{BO}(x) = \underset{y}{\operatorname{argmax}}\ p(y \mid x)$$

So how can we get this Bayes Optimal Classifier?

- $P(y|x)=P(x|y)P(y)/P(x)$
- $P(y|x)=P(y,x)/P(x)$

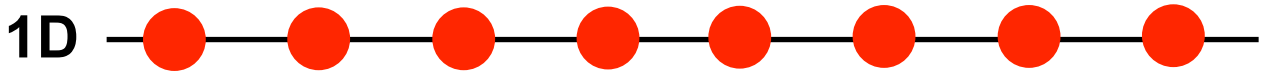
So how can we get this Bayes Optimal Classifier?

- $P(y|x)=P(x|y)P(y)/P(x)$
- $P(y|x)=P(y,x)/P(x)$
- Estimate from the training data.

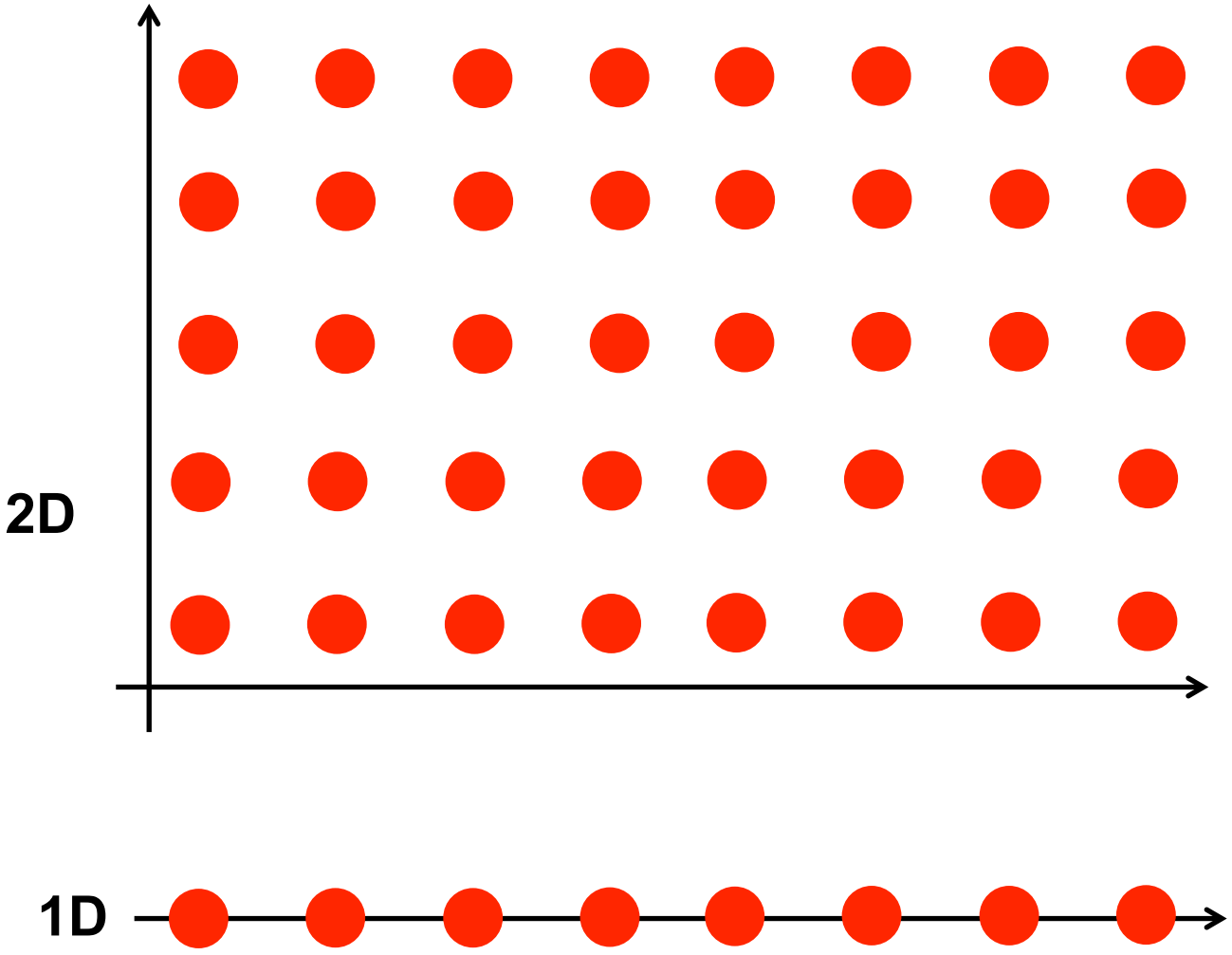
The Law of Large Numbers and Histograms

- The density histogram of *n* samples from a distribution converges to the graph of the underlying probability distribution function of that distribution as $n \to \infty$.

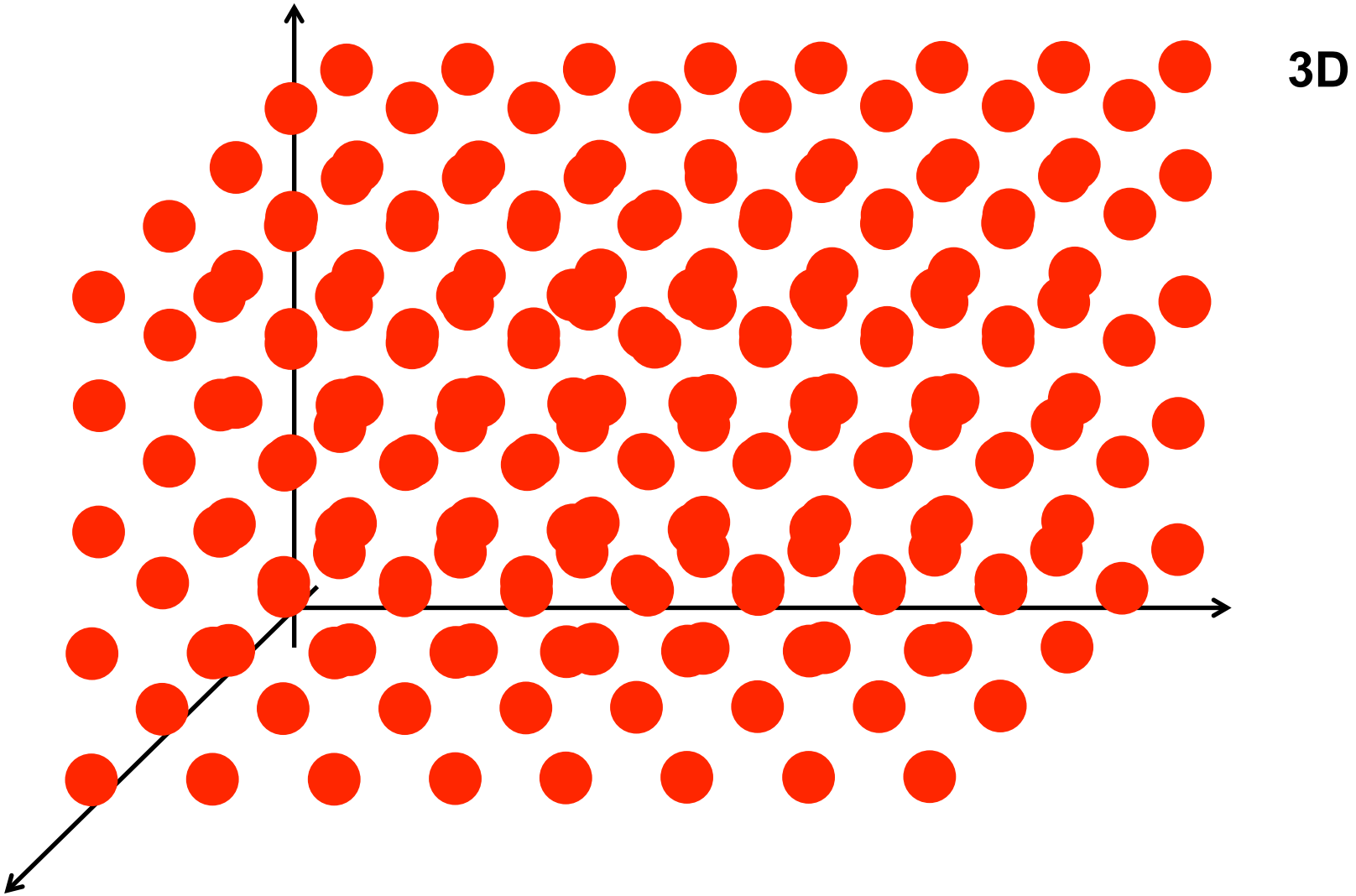
The Curse of Dimensionality



The Curse of Dimensionality



The Curse of Dimensionality



The Curse of Dimensionality