# CSCE 421: Machine Learning

Lecture 1: Introduction to Machine Learning

Texas A&M University
Section 200/500
Bobak Mortazavi

# Welcome to CSCE 421!

- About this class
  - Who are we?
  - Syllabus
  - Texts and other resources
- Intro to machine learning

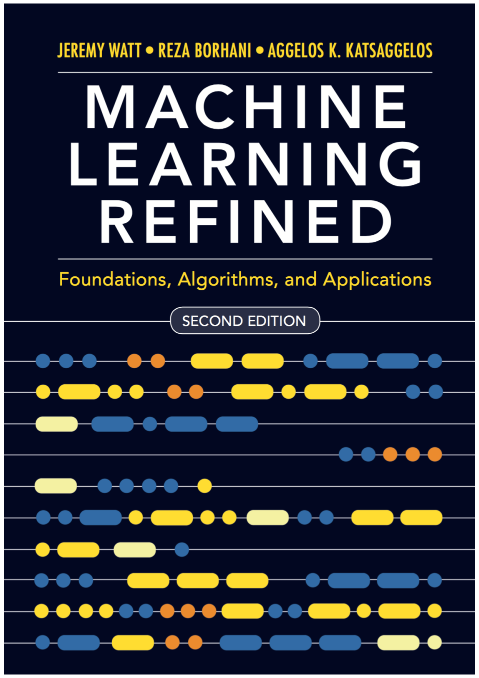
# Who are we?

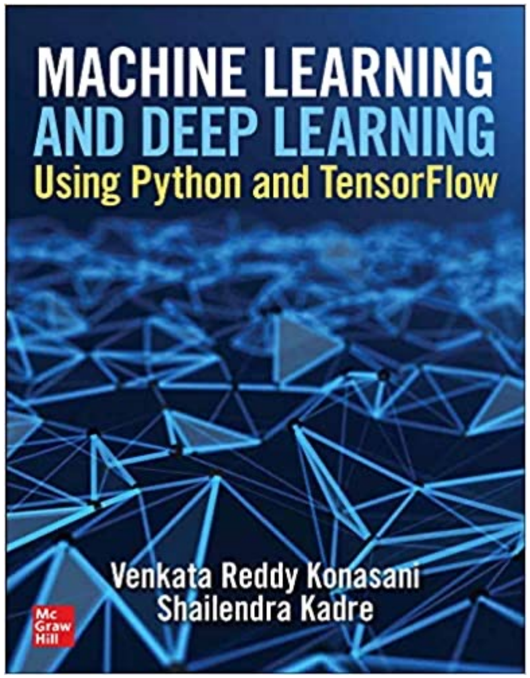
- Instructor
  - Bobak Mortazavi
  - bobakm@tamu.edu – please put [CSCE 421 SP23] in subject line for all emails
- TA
  - Zhale Nowroozilarki
  - zhale@tamu.edu – please put [CSCE 421 SP23] in subject line for all emails
- Grader
  - Dibyanshu Shekhar
  - dshekhar@tamu.edu– please put [CSCE 421 SP23] in subject line for all emails

# Schedule, Syllabus, Canvas

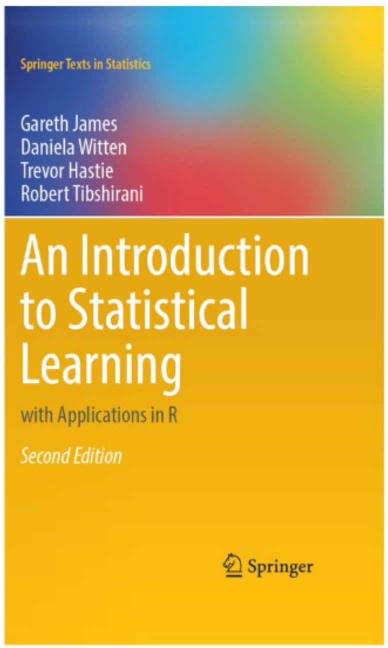
- To Canvas we go!

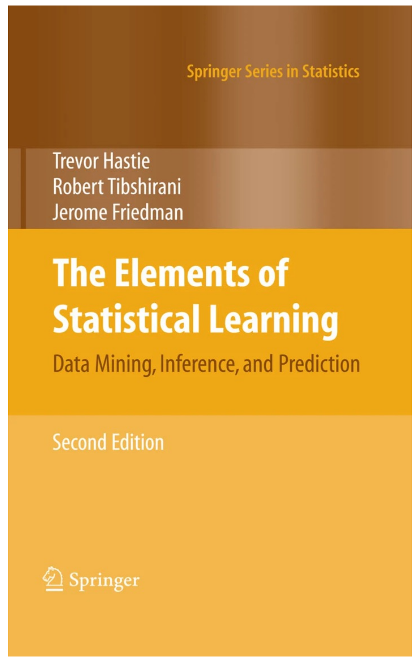
# Text Books

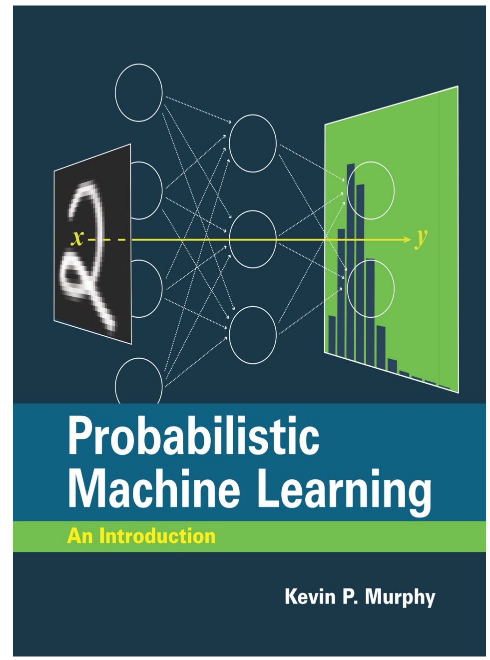

JEREMY WATT • REZA BORHANI • AGGELOS K. KATSAGGELOS
MACHINE LEARNING REFINED
Foundations, Algorithms, and Applications
SECOND EDITION


MACHINE LEARNING
AND DEEP LEARNING
Using Python and TensorFlow
Venkata Reddy Konasani
Shailendra Kadre
Mc Graw Hill

# Other Useful Textbooks



# Required Prerequisites

- The topics you need to know for this class include:
  - Probability
  - Matrix Operations
  - Differentiation
  - Linear Algebra

# Assignments

- Python for programming problems and Latex for written components
- There will be 6 assignments.
- Projects will be individual and may require the development of mdoels with a conference-like write up (more later!)
- Honors section: Will have to propose a final project and have an oral presentation that accompanies report.

# Scribe Notes!

- Counts as part of your HW grade. Due by the final exam time. (no late submissions)
- Have to provide latex write up of lecture notes.
- Need to add additional formulations examples and proofs where applicable and cite additional sources you use (including text books)
- Sign up for a lecture today! (this lecture counts!) No more than 5 students may sign up for each lecture. First come first serve

Email us (MYSELF AND THE TA) the top three choices you would like and we will confirm your selections.

# Late Policy

- Quizzes cannot be made up late. Please check syllabus now and alert us of any conflicts.
- Final project deliverables (due the date/time of the ”final exam” cannot be late.
- No exceptions will be made.
- Materials are due at 11:59:59 pm on the date they are state as due. It is considered late if it arrives at 12:00 am (technically the next day)

# Slip Days

- If an item is due at 11:59:59 pm it is considered late if it arrives at 12:00 am
- For homework assignments, you'll be given 3 slip days to allow flexibility.
- Each slip day gives you 24 additional hours to be used without the assignment being considered late.
- Once you use a slip day it is counted.
- Once you are out of slip days, no late assignments will be allowed.
- At the end of the semester, each slip day you do not use will add 3% to your overall HW grade (allowing for a total possible 109% of homework points)

# Quizzes

- You will take 5 online quizzes (through Canvas) at the start of class on the assigned days.
- Syllabus indicates quiz dates.
- 30 minute quiz.
- You will be graded on your 4 best quizzes.

# Midterms

- There will be two midterm examinations (see syllabus for assigned date)
- Midterms cannot be made up later
- There is no final exam

# Final Projects

- More details later but – will be due the day of the final exam (end of day)
- Will require code + written report
- Honors section will have to propose their own project and turn in an oral presentation as well
- All are welcome to propose a project if it aligns with their research
- **REGULAR SECTION:** Top 5 scoring groups (limited to size of 3) will have lowest quiz score dropped. HONORS SECTION NOT ELIGIBLE FOR THIS.

# Welcome to CSCE 421!

- About this class
  - Who are we?
  - Syllabus
  - Texts and other resources
- Intro to machine learning

# Acknowledgements

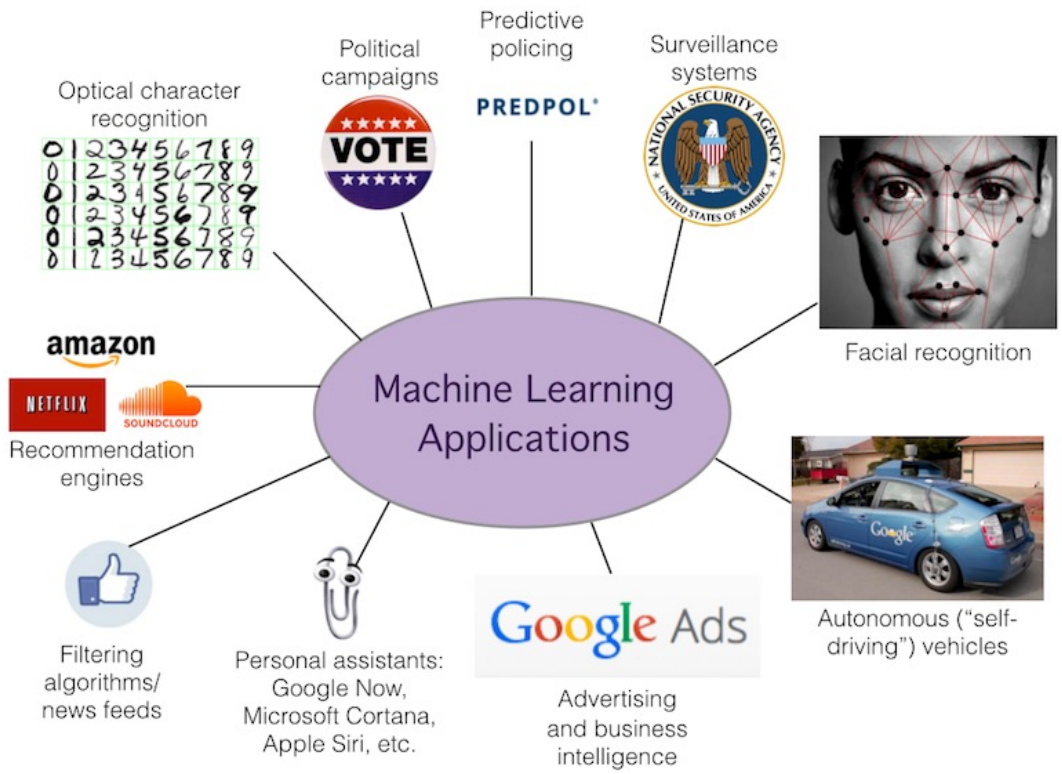

- Material from textbooks used in slides from:
  - Watt, Borhani, and Katsaggelos
  - Friedmen, Hastie, and Tibshirani
  - James, Witten, Hastie, and Tibshirani
  - Examples and Images Used

# Topics

- About This Class
- Introduction to Machine Learning
  - What is machine learning?
  - Models and accuracy
  - Some general supervised learning examples

# What is machine learning?

# What is machine learning?



# Let’s start with an example

- Let’s say you want to teach a computer how to tell the difference between a cat and dog.

# How do you describe the differences?

- Let’s say you want to teach a computer how to tell the difference between a cat and dog

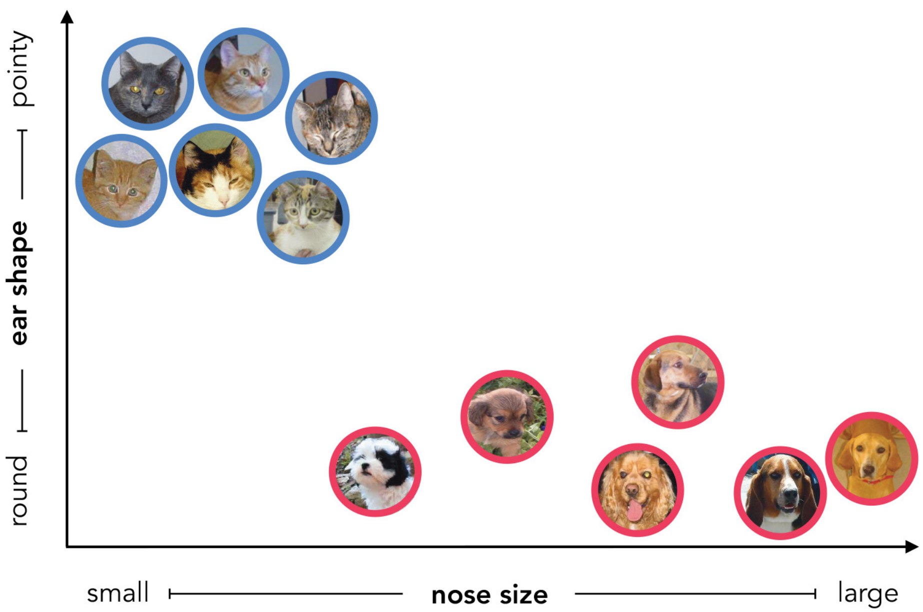
- Assume you take these pictures (call this data collection)
- Take a moment and ask yourself – how would you tell the difference between a cat and dog?

# Feature extraction

- Let’s say you want to teach a computer how to tell the difference between a cat and dog.

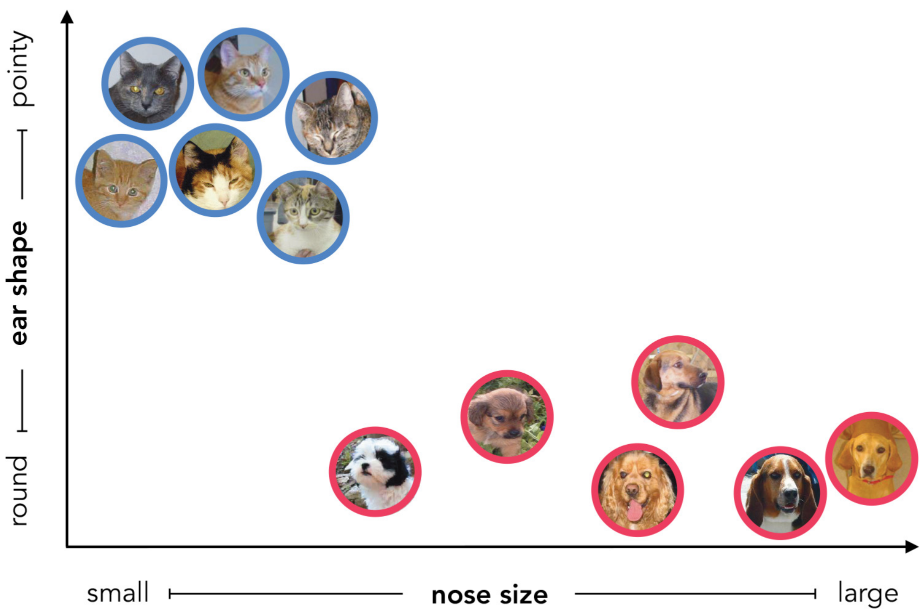
- Those things you just described? Those can be considered features (or covariates)
- So now we have data collection and feature design (or feature engineering) as two important steps.

# Picking good features!



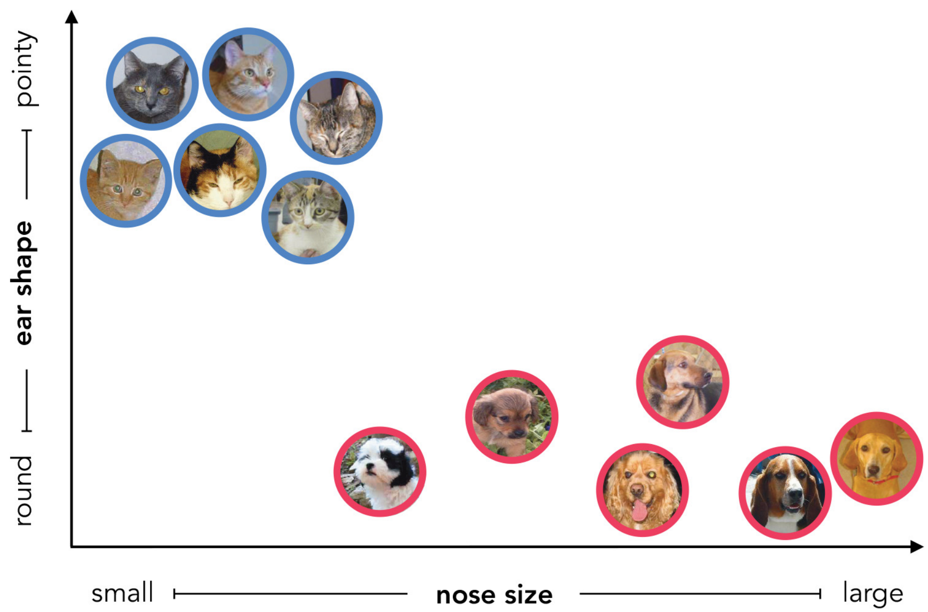

- Let's see how well some features distinguish
- Example: *size of nose* and *shape of ear*

# Picking good features!



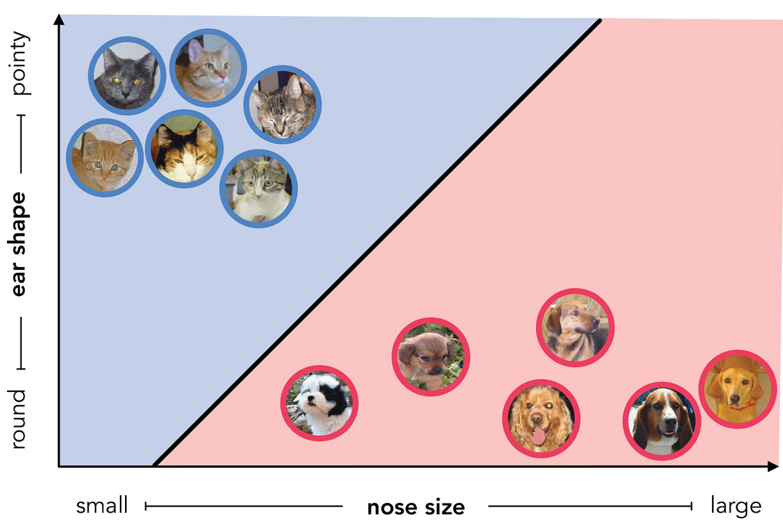

- Let's see how well some features distinguish
- Example: *size of nose* and *shape of ear*
- These seem to separate (or discriminate) between cats and dogs well!

# Picking good features!



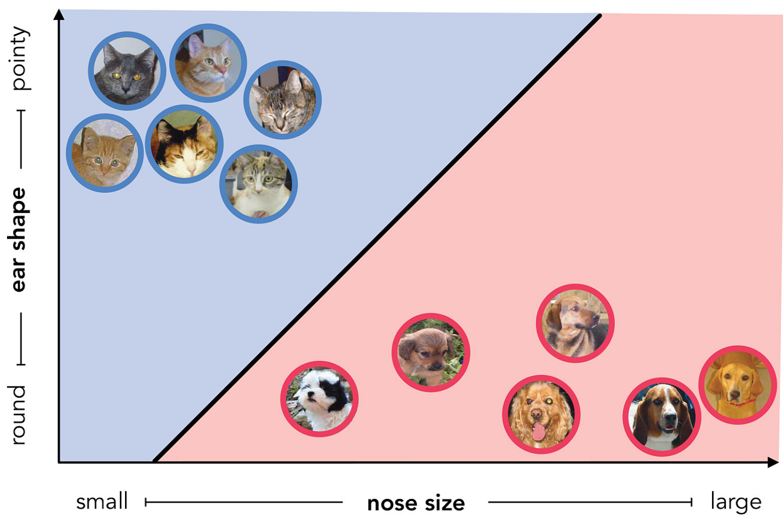

- Let’s see how well some features distinguish
- Example: *size of nose* and *shape of ear*
- These seem to separate (or discriminate) between cats and dogs well!
- Can we mathematically describe this difference?

# Designing a Model




- This line seems to separate the space where cats and dogs are represented
- How would you describe this line that separates these objects?

# Designing a Model




- This line seems to separate the space where cats and dogs are represented
- The line is described by a Slope + Intercept. Finding these terms is done via different mathematical optimization techniques.

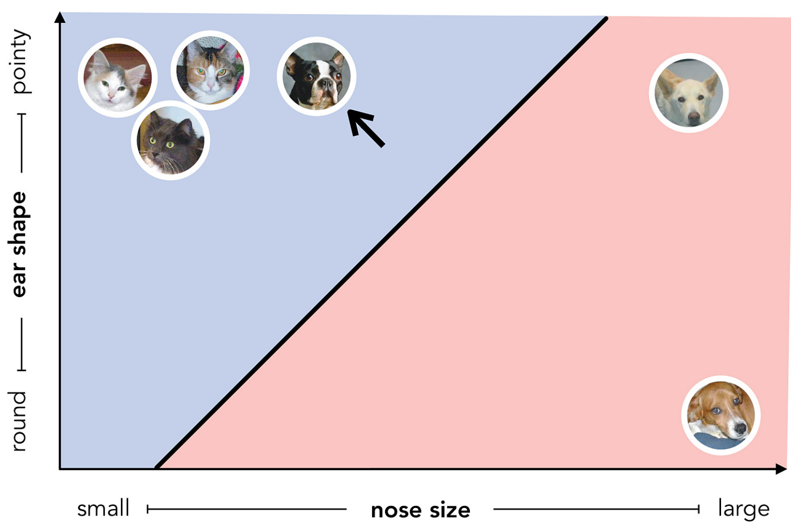
# Testing your model

- Are these dogs or cats?

# Testing your model

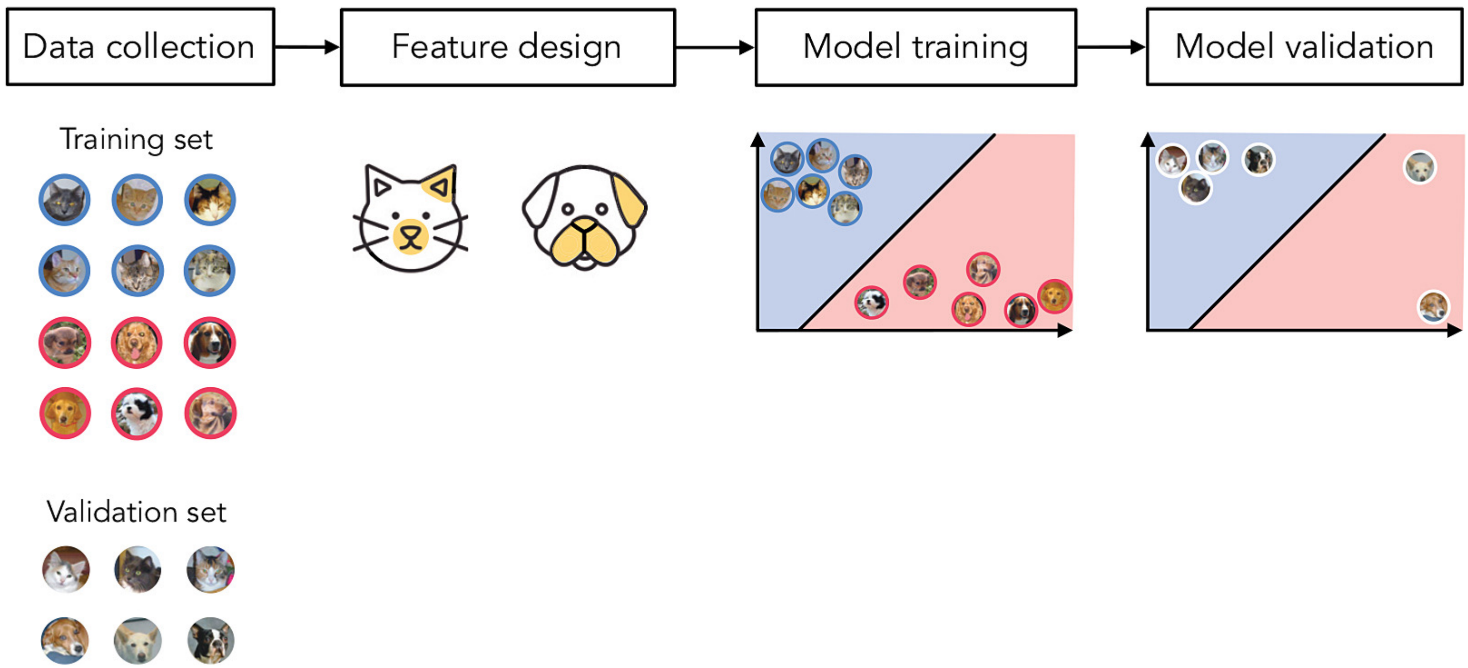
- Are these dogs or cats?
- How would you validate your model?
- What ways would you describe if your model is good?

# Evaluating Test Cases

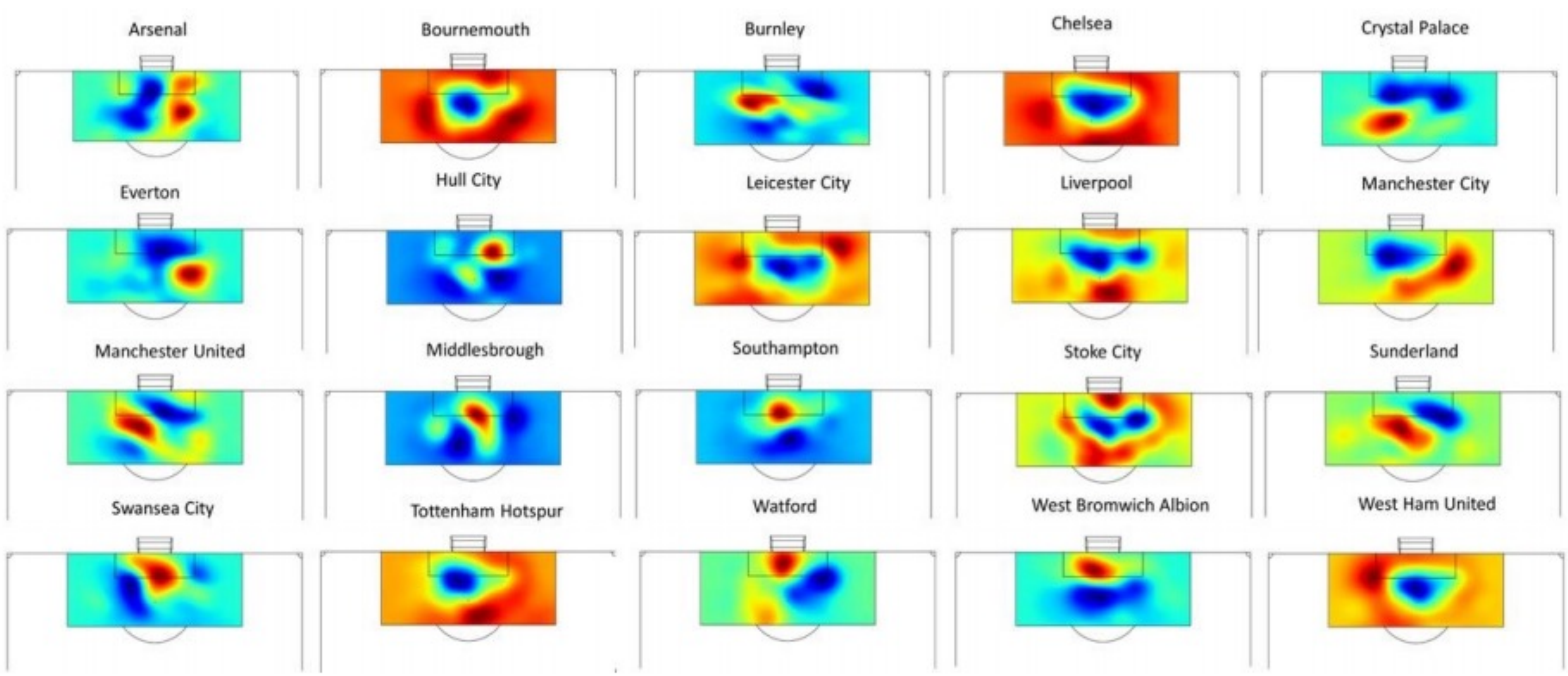



- After you build a model you must validate that it works on new, unseen data!
- Might not work as well as the training version

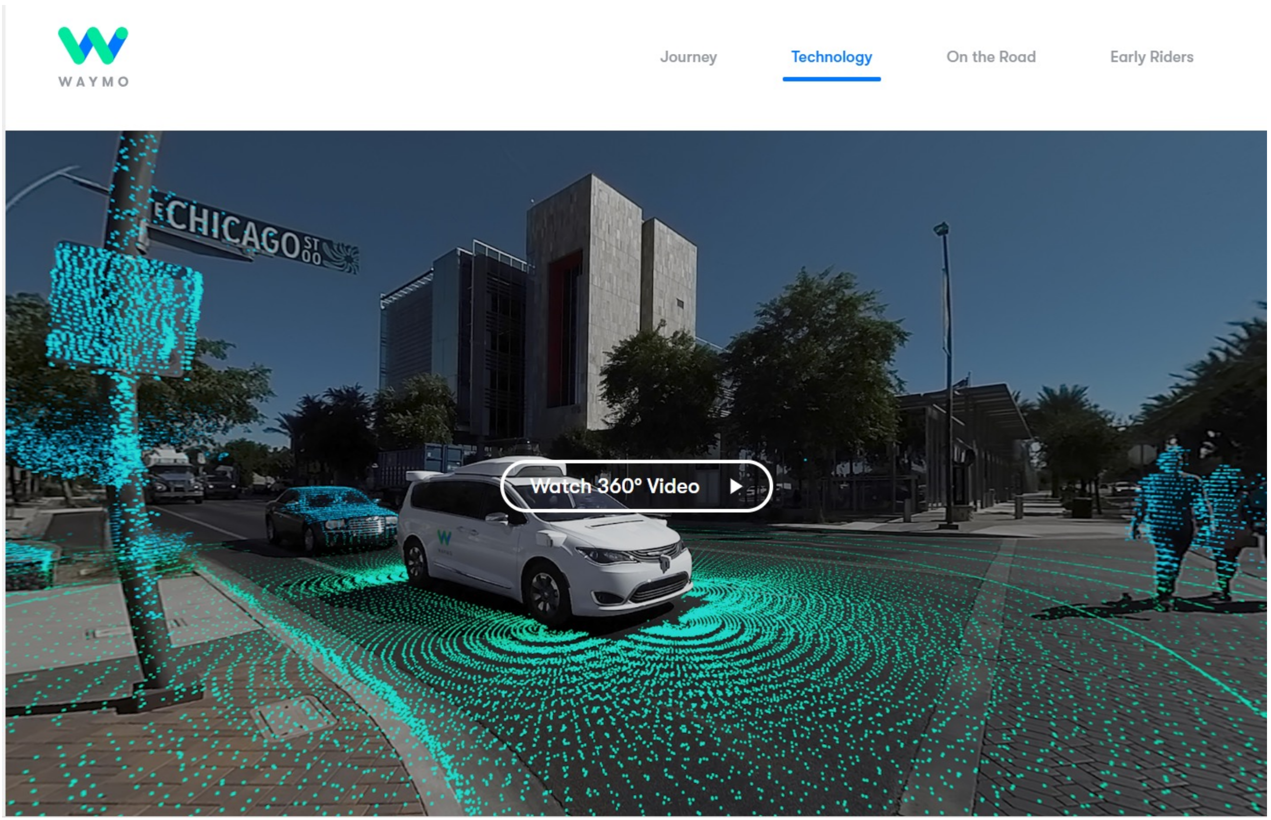
# General Machine Learning Pipeline



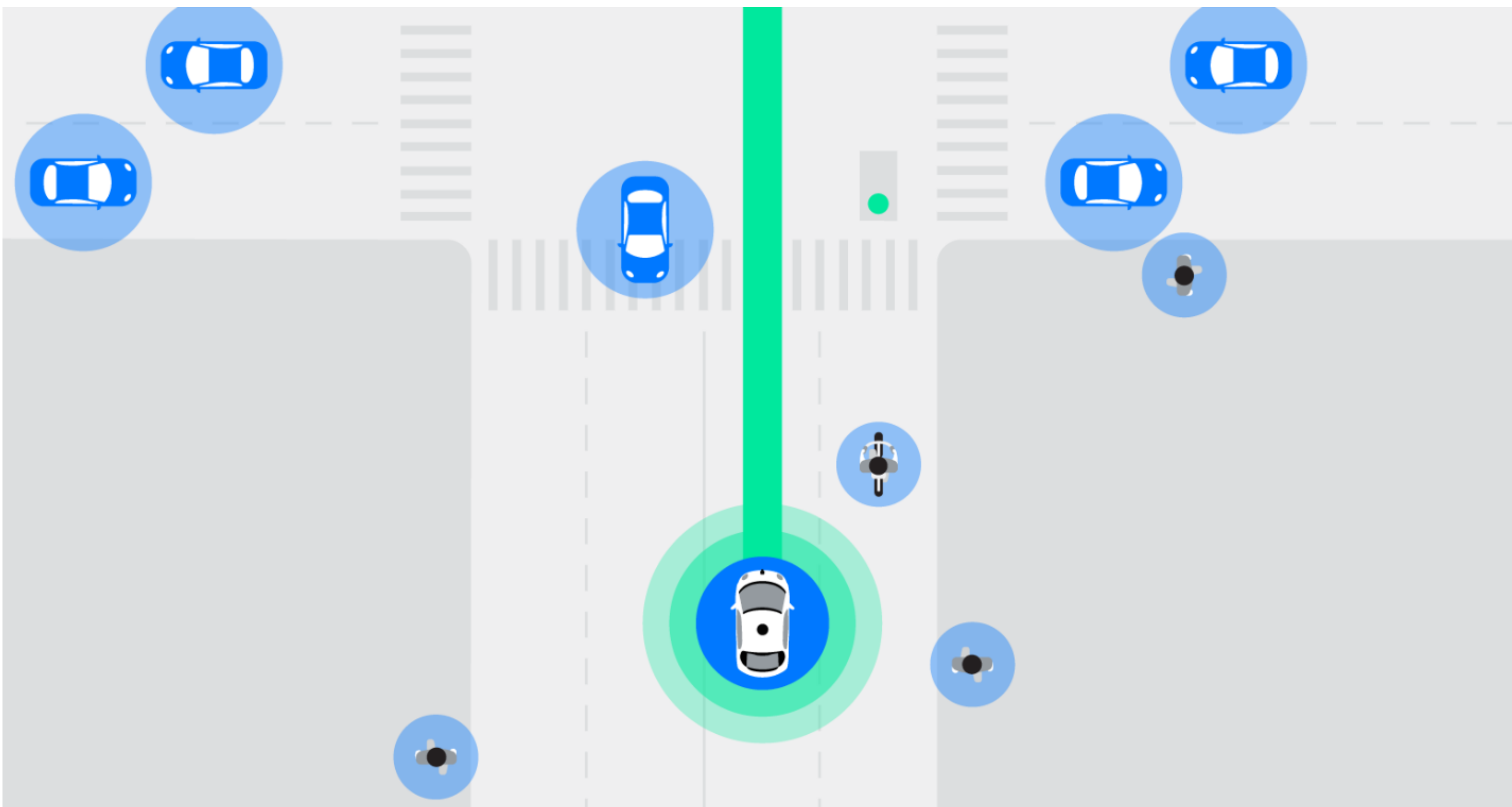
# Some other examples of machine learning!



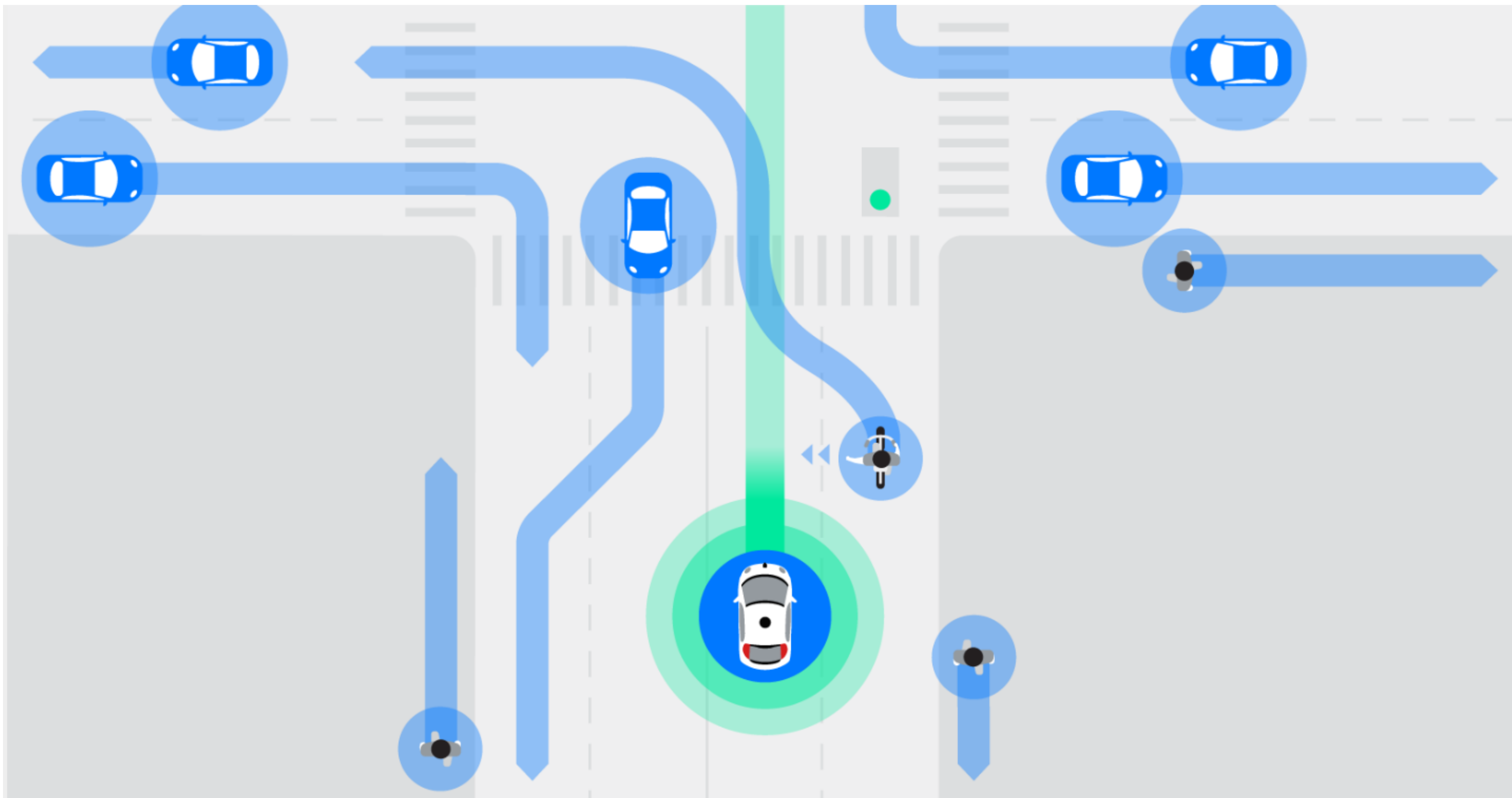
# Waymo.com/tech



# Self Driving Challenges



# Self Driving Challenges



# Advertising Example

- Assume we are trying to improve sales of a product through advertising in three different markets:

  TV Newspaper Online/Mobile
- Then we can define the advertising budget as:
  - Input variables $x_i$: Consisting of TV budget $x_{i,1}$, newspaper budget $x_{i,2}$, and mobile budget $x_{i,3}$ (also known as the independent variables)
  - Output variable $y$: sales numbers (also known as the response variable or dependent variable)
  - Can generalize $x_{i,1}, x_{i,2}, \ldots, x_{i,m} = x_i$

# Advertising Example

- Assume we are trying to improve sales of a product through advertising in three different markets:

  TV        Newspaper        Online/Mobile

- Then we can define the advertising budget as:
  - Input variables $x_i$: Consisting of TV budget $x_{i,1}$, newspaper budget $x_{i,2}$, and mobile budget $x_{i,3}$ (also known as the independent variables)
  - Output variable $y$: sales numbers (also known as the response variable or dependent variable)
  - Can generalize $x_{i,1}, x_{i,2}, \ldots, x_{i,m} = x_i$
  - We then seek to model the relationship $Y = f(X) + \varepsilon$ where f is the systemic information and $\varepsilon$ is random error

# Advertising Example

- Assume we are trying to improve sales of a product through advertising in three different markets:

  TV Newspaper Online/Mobile

- Then we can define the advertising budget as:
  - Input variables $x_i$: Consisting of TV budget $x_{i,1}$, newspaper budget $x_{i,2}$, and mobile budget $x_{i,3}$ (also known as the independent variables)
  - Output variable $y$: sales numbers (also known as the response variable or dependent variable)
  - Can generalize $x_{i,1}, x_{i,2}, \ldots, x_{i,m} = x_i$
  - We then seek to model the relationship $Y = f(X) + \varepsilon$ where f is the systemic information and $\varepsilon$ is random error
  - How do we learn f?

# Why estimate f? (and how?)

- Prediction

- Inference

# Why estimate f? (and how?)

- Prediction

  In many situations, $x_i$ is available to us but y is not. We would like to estimate y as accurately as possible

- Inference

# Why estimate f? (and how?)

- Prediction

  In many situations, $x_i$ is available to us but y is not. We would like to estimate y as accurately as possible

- Inference

  In some situations both $x_i$ and y are available and we would like to understand how y is affected by changes in $x_i$

# Predicting f?

- Seek to learn an $\hat{f}(x)$ that generates a $\hat{y}$, which estimates y as closely as possible by attempting to minimize the reducible error.
- We have to accept that there may always be irreducible error that is independent of x. Why might that be?

# Predicting f?

- Seek to learn an $\hat{f}(x)$ that generates a $\hat{y}$, which estimates y as closely as possible by attempting to minimize the reducible error.
- We have to accept that there may always be irreducible error that is independent of x. Why might that be?
- We can calculate the expected value of the average error as:

$$\mathbb{E}\,(y - \hat{y}) = \mathbb{E}[f(x) + \varepsilon - \hat{f}(x)]^2 = [f(x) - \hat{f}(x)]^2 + Var(\varepsilon)$$

(Scribe notes – please provide entire derivation)

- The focus of this class is to learn the various techniques that estimate f and minimize this reducible error.

# So how do we estimate f?

- We need training data to learn from

# So how do we estimate f?

- We need training data to learn from
- Define training samples $S = (x_1, y_1), \dots, (x_n, y_n)$ where

$$x_i = (x_{i,1}, x_{i,2}, \dots x_{i,p})^T$$

for n subjects and p dimensions.

# So how do we estimate f?

- We need training data to learn from
- Define training samples $S = (x_1, y_1), \dots, (x_n, y_n)$ where
$$x_i = (x_{i,1}, x_{i,2}, \dots x_{i,p})^T$$
  for n subjects and p dimensions.
- Where y can be:
  - Categorical or nominal – the problem we solve is classification or pattern recognition
  - Continuous – the problem we solve is regression
  - Categorical with order, or ordinal (example: Grades A, B, C, D, F) – the problem we can solve is ordinal regression

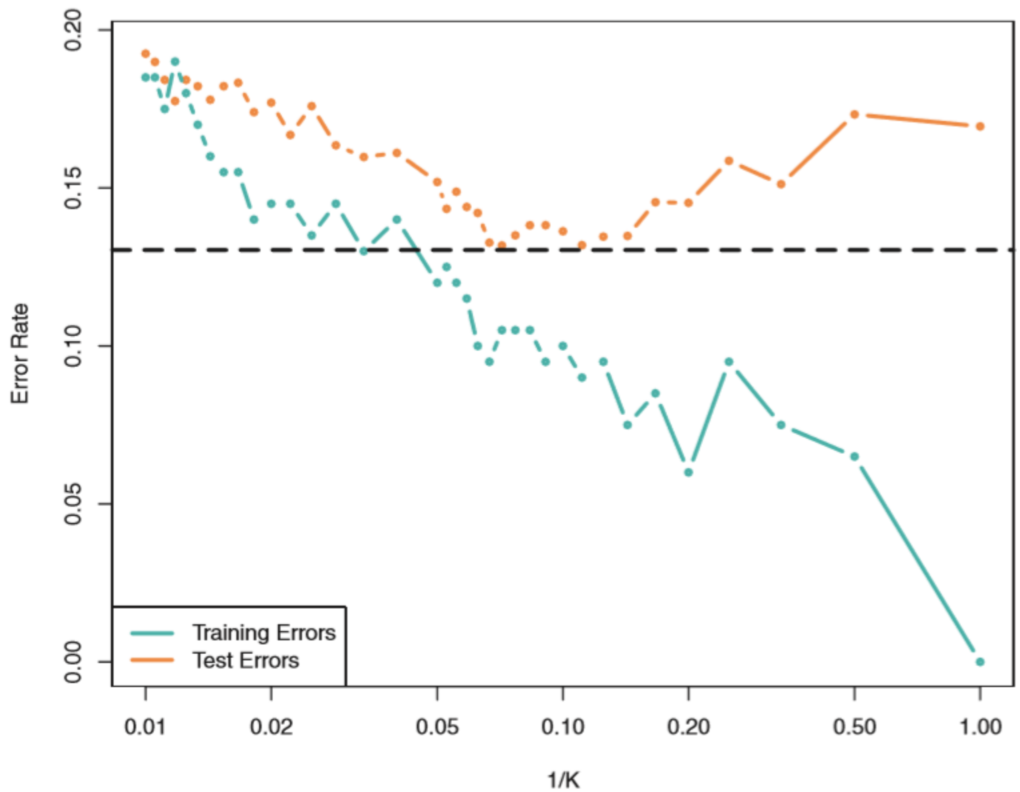
# So how do we estimate f?

- We need training data to learn from
- Define training samples $S = (x_1, y_1), \dots, (x_n, y_n)$ where

$$x_i = (x_{i,1}, x_{i,2}, \dots x_{i,d})^T$$

  for n subjects and p dimensions.
- Where y can be:
  - Categorical or nominal – the problem we solve is classification or pattern recognition
  - Continuous – the problem we solve is regression
  - Categorical with order, or ordinal (example: Grades A, B, C, D, F) – the problem we can solve is ordinal regression
- Please Note! NOTATIONS MAY DIFFER ACROSS BOOKS

# Things to watch out for!



# Standard Machine Learning Terms

## Tasks

- Machine Learning tasks typically fall into the following tasks:
  - Classification - assigning a category to each item i.e. text classification, speech recognition, cats vs. dogs
  - Regression - learning a real value for each item i.e. stock value prediction
  - Clustering – learning how to partition a set of items into homogenous subsets i.e. identifying communities within larger groups of people
  - Dimension Reduction - learning a transformation from the initial representation into a lower-dimensional representation i.e. digital image processing

## Terminology

# Standard Machine Learning Terms

## Tasks

- Machine Learning tasks typically fall into the following tasks:
  - Classification - assigning a category to each item i.e. text classification, speech recognition, cats vs. dogs
  - Regression - learning a real value for each item i.e. stock value prediction
  - Clustering – learning how to partition a set of items into homogenous subsets i.e. identifying communities within larger groups of people
  - Dimension Reduction - learning a transformation from the initial representation into a lower-dimensional representation i.e. digital image processing

## Terminology

- Key terms and definitions common in machine learning
  - Examples - instances of data for learning or evaluation
  - Features – set of attributes representing the examples (often a vector)
  - Labels - values/categories assigned to the examples
  - Hyperparameters – free parameters of a model not determined in training but rather specific as inputs
  - Training sample – examples used in the training process
  - Validation sample – examples used in parameter tuning
  - Test sample - examples used for performance evaluation
  - Loss function – a function measuring the difference between a predicted label and the true label. This function is to be minimized in the optimization stage
  - Hypothesis set – a set of possible functions which map the feature vectors to the set of labels

# Other Learning Scenarios

- Online Learning – training and testing phases are intermixed. The data is made available to the learner over time. The objective is to minimize the cumulative loss over all rounds.
- Reinforcement Learning – training and testing phases are intermixed. The learner interacts with/affects the environment. Some actions of the learner lead to being rewarded. Objective is to maximize the rewards based on a course of actions. Exploration vs. exploitation dilemma.
- Self-Supervised Learning – the learner interactively collects training examples, querying an oracle for requesting new labels. Used when labels are expensive/difficult to obtain.

# Takeaways and Next Time

- Course structure
- What are machine learning tasks?
- What are some basics of machine learning?
- Can a learner always achieve zero prediction error?
- Next session: Review of Linear Algebra, Measures of Model Performance, Linear Regression and K-Means