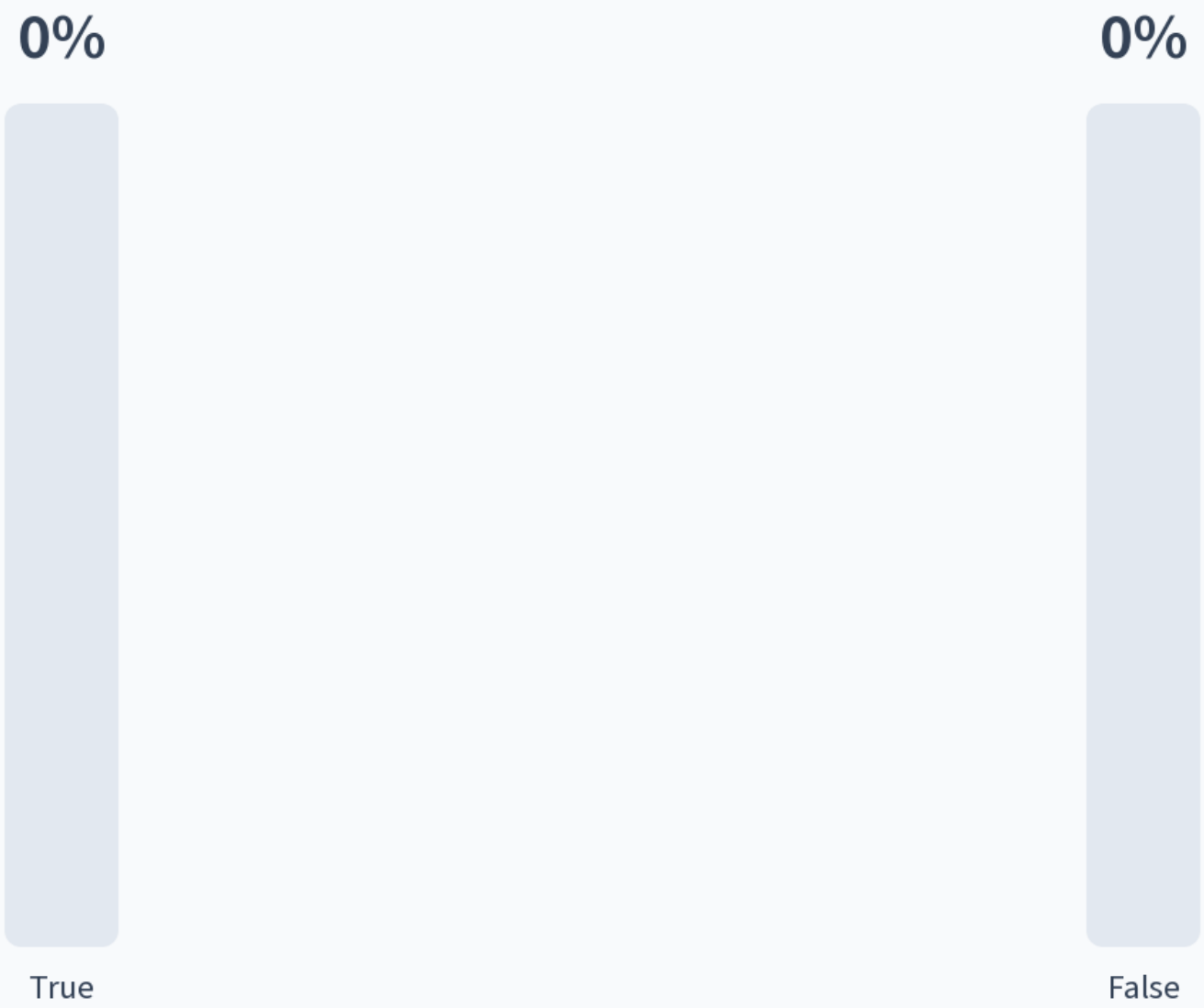
MACHINE LEARNING REVIEW

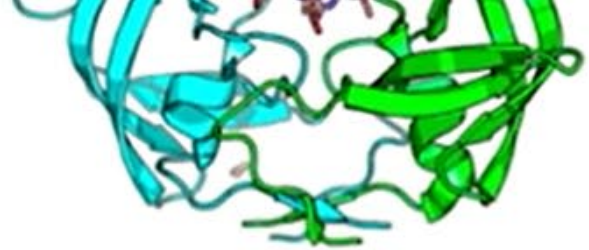
DATA3320

Applications of machine learning are all around us。

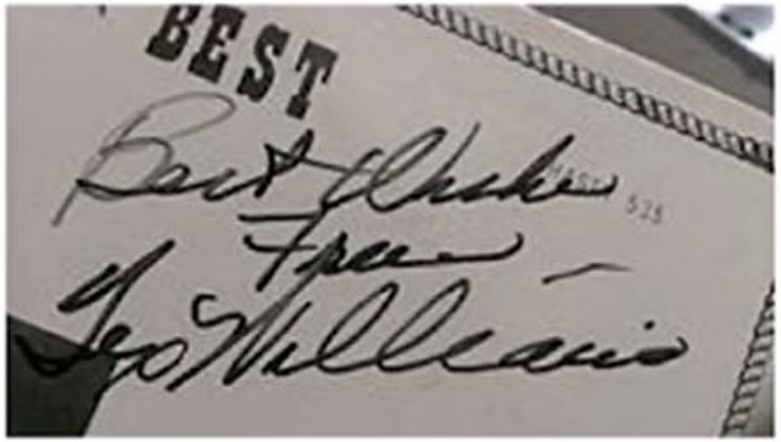


AlphaGo

Recommendation systems



Drug discovery



Character recognition

2σ TWO SIGMA

Hedge fund stock predictions

Voice assistants

Assisted driving

Face detection/recognition

Cancer diagnosis

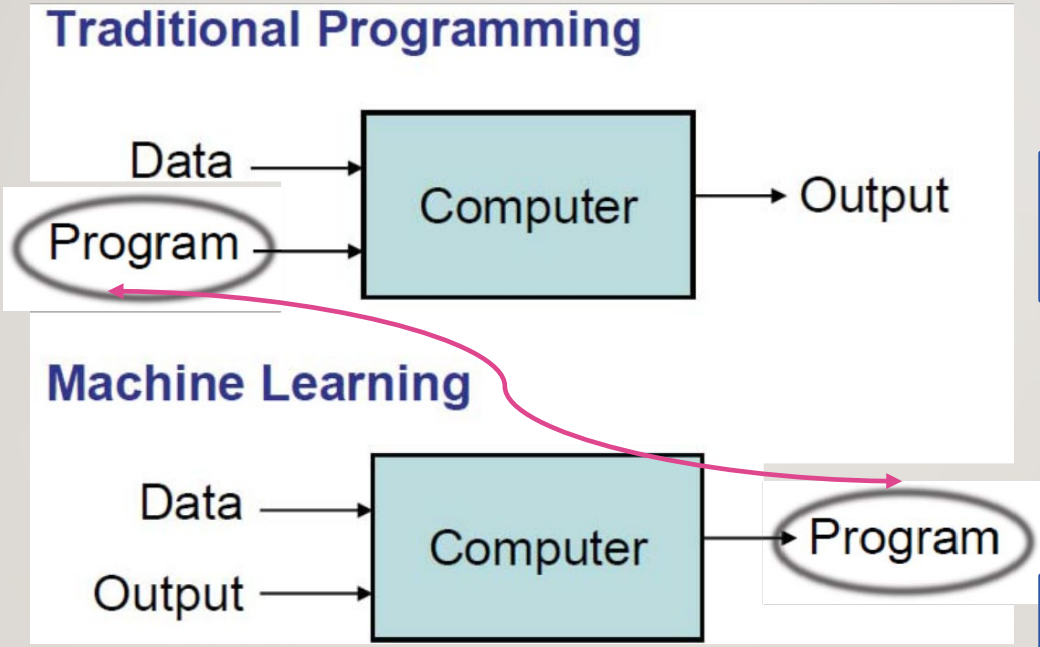
WHAT IS MACHINE LEARNING?

- All useful programs "learn" something
- Early definition of machine learning:

  "Field of study that gives computers the ability to learn without being explicitly programmed." Arthur Samuel (1959)

  - Computer pioneer who wrote first self-learning program, which played checkers - learned from "experience"
  - Invented alpha-beta pruning - widely used in decision tree searching

WHAT IS MACHINE LEARNING?



Square root finder

Curve fitting by linear regression

WHAT'S THE KEY?

- We don't tell the computer how to make predictions.
- We give the computer old examples, it figures out how to make predictions itself.

OLD CONCEPTS, NEW APPLICATIONS

- That idea isn't new (how do you predict if a restaurant will be good? If a job is worth applying for? If someone will date you?)
- The scale, breadth, speed, and accuracy are new.

What is the technical difference between classical ML and deep learning?

Classical ML was invented first.

0%

The use of neural networks.

0%

Deep learning is used in robots.

0%

None of the above

0%

HOW ARE THINGS LEARNED?

Interested in extending to programs that can infer useful information from **implicit** patterns in data

- Memorization
  - ❑ Accumulation of individual facts

  Limited by
  - ❖ Time to observe facts
  - ❖ Memory to store facts

Declarative knowledge

- Generalization
  - ❑ Deduce new facts from old facts

  Limited by accuracy of deduction process
  - ❖ Essentially a predictive activity
  - ❖ Assumes that the past predicts the future

Imperative knowledge

BASIC PARADIGM

- Observe set of examples: training data
- Infer something about process that generated that data
- Use inference to make predictions about previously unseen data: test data

Football players, labeled by position, with height and weight data

Find canonical model of position, by statistics

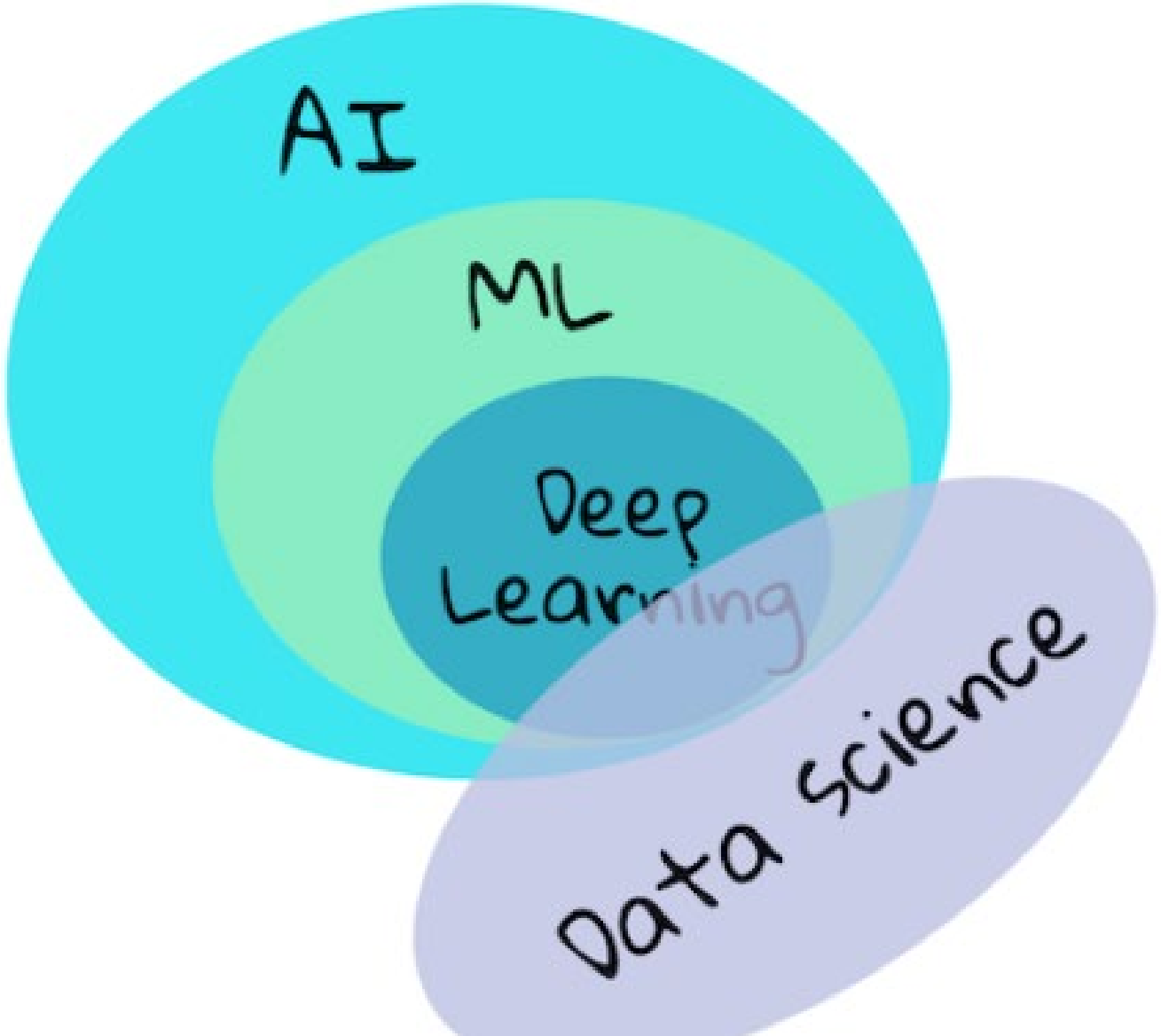
Predict position of new players

- Variations on paradigm
  - Supervised: given a set of feature/label pairs, find a rule that predicts the label associated with a previously unseen input
  - Unsupervised: given a set of feature vectors (without labels} group them into "natural clusters" (or create labels for groups}

WHAT'S THIS DEEP LEARNING BIT?

- Deep Learning uses neural networks – one type of algorithm of machine learning.
- The DL algorithm can be replaced with linear regression, trees, or any other algorithm and the concepts are the same.

AI, ML, Deep Learning



WHY MACHINE LEARNING?

✅ Think for a minute why a business would want to try to use machine learning strategies vs. creating a hard-coded rules-based engine.

MACHINE LEARNING LINGO

What?	Parameters	Structure	Hidden concepts	
What from?	Supervised	Unsupervised	Reinforcement	Self-supervised
What for?	Prediction	Diagnosis	Compression	Discovery
How?	Passive	Active	Online	Offline
Output?	Classification	Regression	Clustering	
Details??	Generative	Discriminative	Smoothing	

Classical Machine Learning

Task Driven

Data Driven

Supervised Learning

(Pre Categorized Data)

Classification

(Divide the socks by Color)

Eg. Identity Fraud Detection

Regression

(Divide the Ties by Length)

Eg. Market Forecasting

Unsupervised Learning

(Unlabelled Data)

Clustering

(Divide by Similarity)

Eg. Targeted Marketing

Association

(Identify Sequences)

Eg. Customer Recommendation

Dimensionality Reduction

(Wider Dependencies)

Eg. Big Data Visualization

Obj: Predications & Predictive Models

Pattern/ Structure Recognition

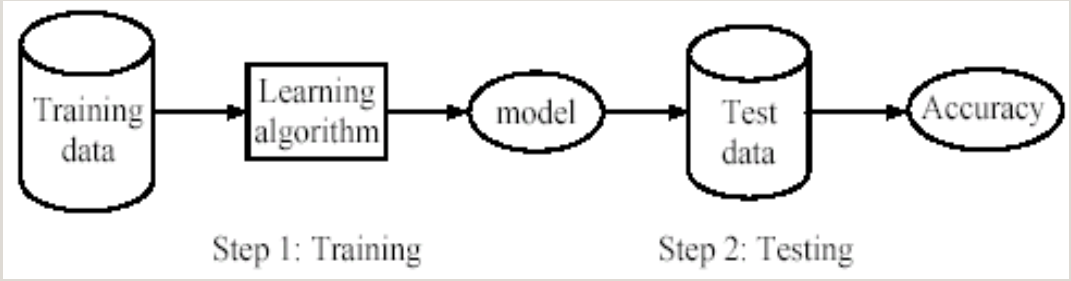
THE PROCESS IS SIMPLE

- 1 – Collect data on what we care about.
- 2 – ‘Clean’ that data up.
- 3 – Give the data to the algorithm to learn.
- 4 – Let the model generated from the algorithm make new predictions.

SUPERVISED LEARNING PROCESS: TWO STEPS

- Learning (training): Learn a model using the training data
- Testing: Test the model using unseen test data to assess the model accuracy

$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$



WHAT DO WE MEAN BY LEARNING?

- Given
  - a data set *D*,
  - a task *T*, and
  - a performance measure *M*,

  a computer system is said to **learn** from *D* to perform the task *T* if after learning the system's performance on *T* improves as measured by *M*.
- In other words, the learned model helps the system to perform *T* better as compared to no learning.

SUPERVISED LEARNING

enjoy algorithms

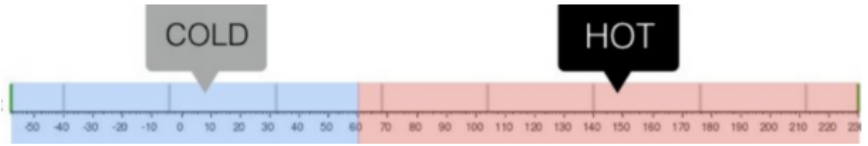
Regression

What will be the temperature tomorrow?

Fahrenheit

Classification

Will it be hot or cold tomorrow?



Fahrenheit

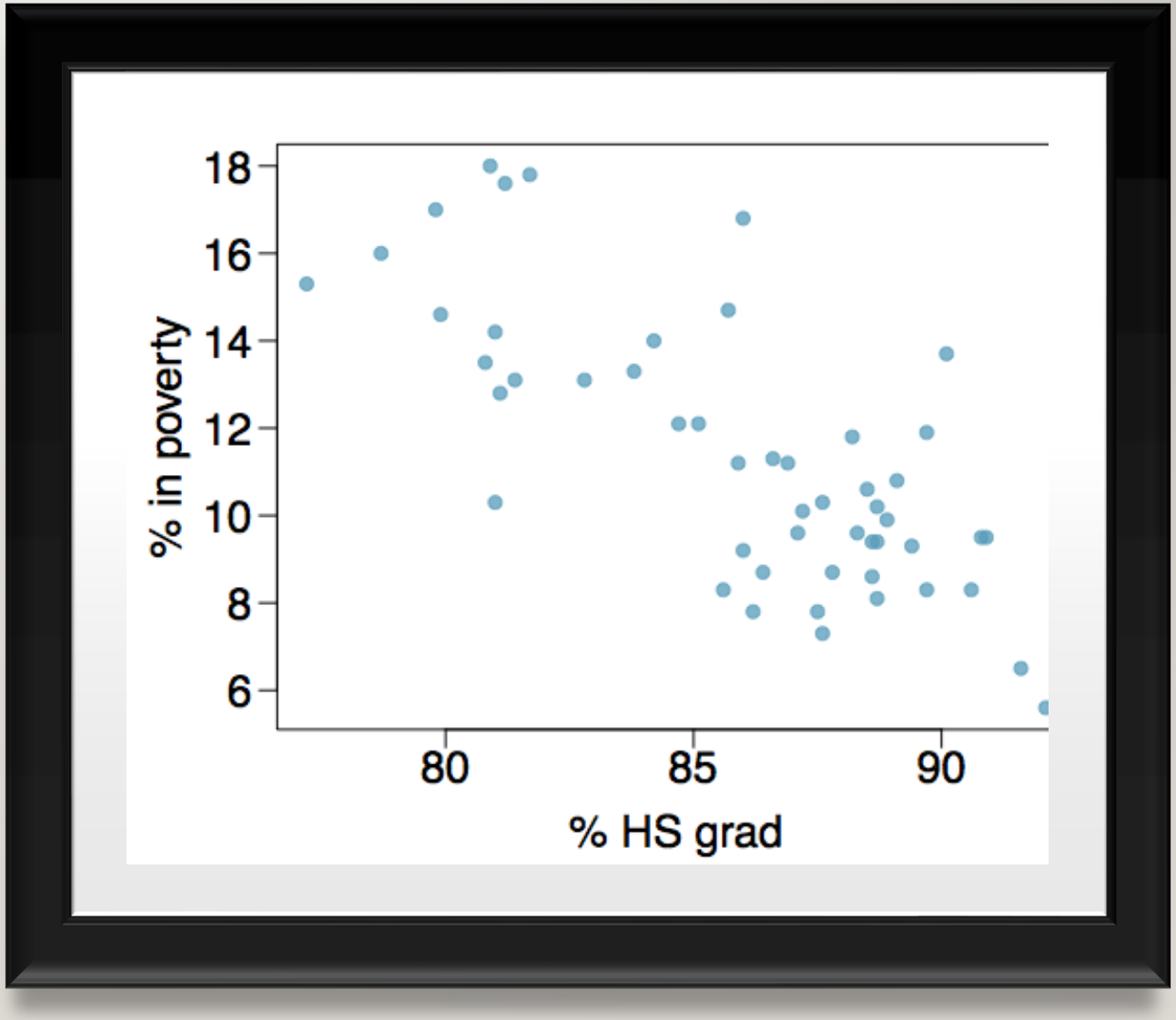
REGRESSION

LINEAR REGRESSION AND PREDICTIVE MODELS

- Most people have done linear regression before – it is making a line of best fit.
  - Result – y = m*x + b line.
  - M = slope, b = y intercept.
- **This process is also a simple predictive model – we provide X and get a prediction for Y.**
- The "regression calculation" uses the training data to "learn" how to generate Y from X.
  - That's the machine learning bit.
- The process of creating a regression is almost the same as other models, we'll do later.

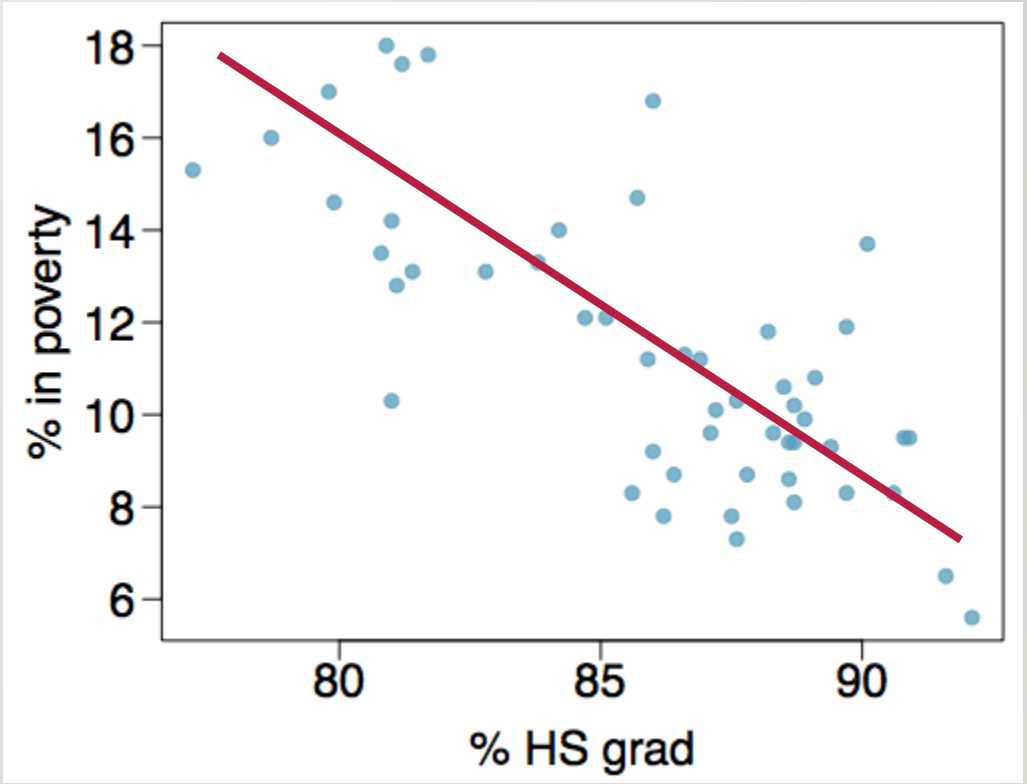
POVERTY VS. HS GRADUATE RATE

- The *scatterplot* below shows the relationship between HS graduate rate in all 50 US states and DC and the percent of residents who live below the poverty line
(income below $23,050 for a family of 4 in 2012).



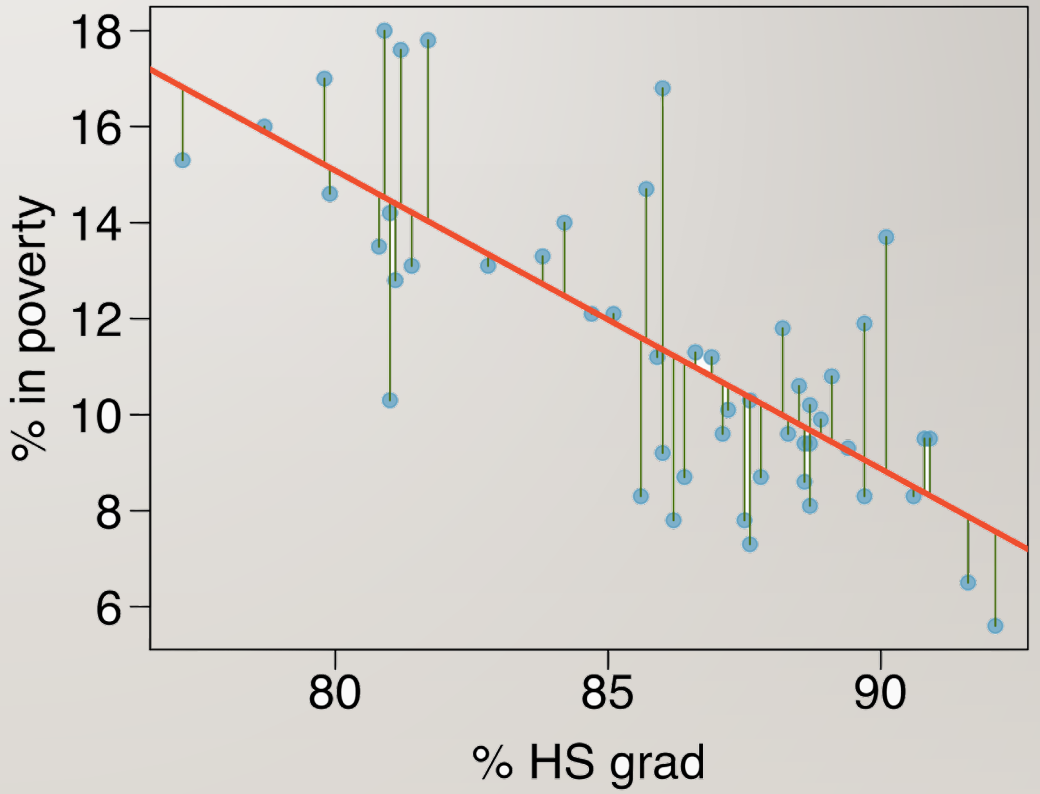
POVERTY VS. HS GRADUATE RATE

We could draw a line of best fit… but how do we know exactly where it goes?



RESIDUALS

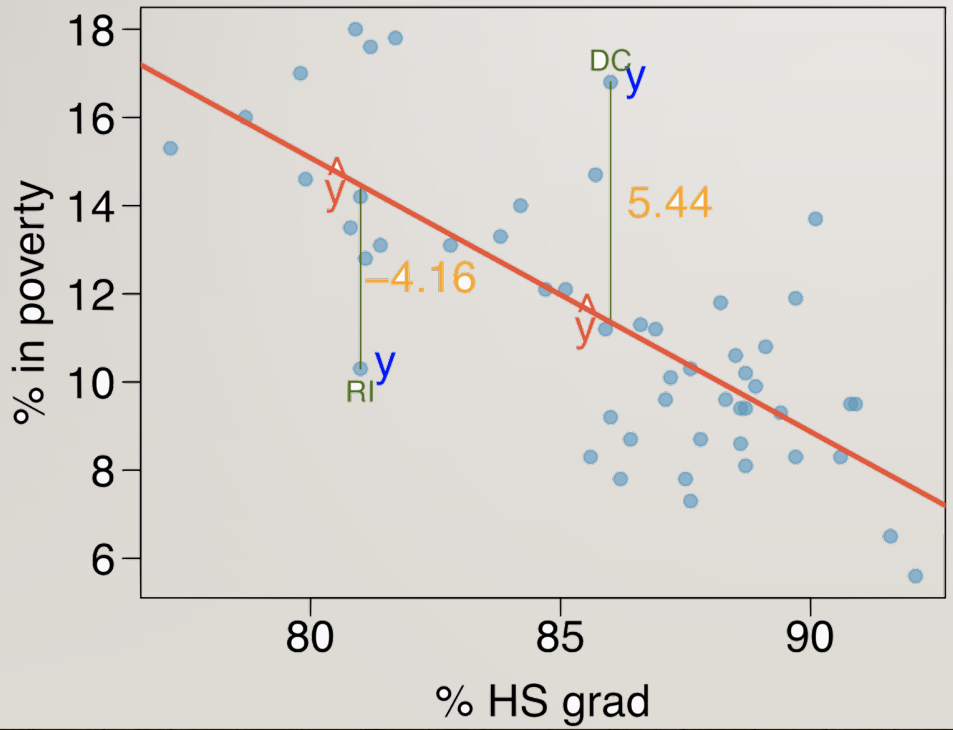
- **Residuals** are the leftovers from the model fit:
  - Data = Fit + Residual



RESIDUALS (CONT.)

Residual is the difference between the observed (y_i) and predicted $\hat{y}_i$.

$$e_i = y_i - \hat{y}_i$$



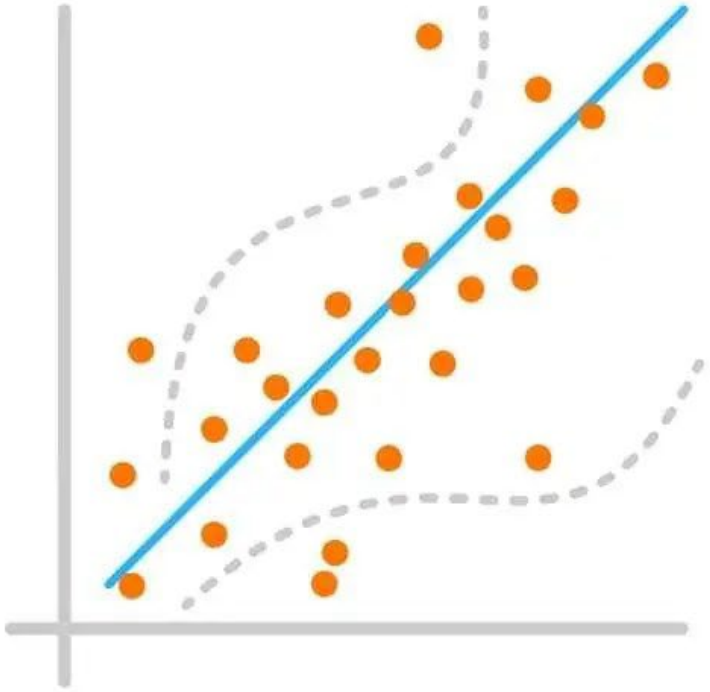
% living in poverty in DC is 5.44% more than predicted.

% living in poverty in RI is 4.16% less than predicted.

OTHER PACKAGES

- Linear regression is performed by many existing packages, such as StatsModels, Scipy, and Scikitlearn.
- The book uses StatsModels when multiple regression starts.
- Which you use mostly doesn't matter, it is a personal choice.
- We'll use both StatsModels and Scikitlearn:
  - Statsmodels provide more stats data in the output, so we will use that sometimes.
  - The scikitlearn is probably more relevant experience for ML stuff.
- I think going forward I might replace some of the statsmodels examples in future workbooks with sklearn one. The interface is easier, and it is more relevant to ML.

4 Common Regression Metrics



1. Mean Squared Error (MSE)
2. Root Mean Squared Error (RMSE)
3. Mean Absolute Error (MAE)
4. R-squared (R^2)

CONCLUSION

- We can train our models to predict Y, given X.
  - In this case, the model is a simple algebra equation.
- This is a simple version of all the more complex ML work to come later.
- The residuals give us information on how good our model is.

- Accuracy and reliability of the predictions.

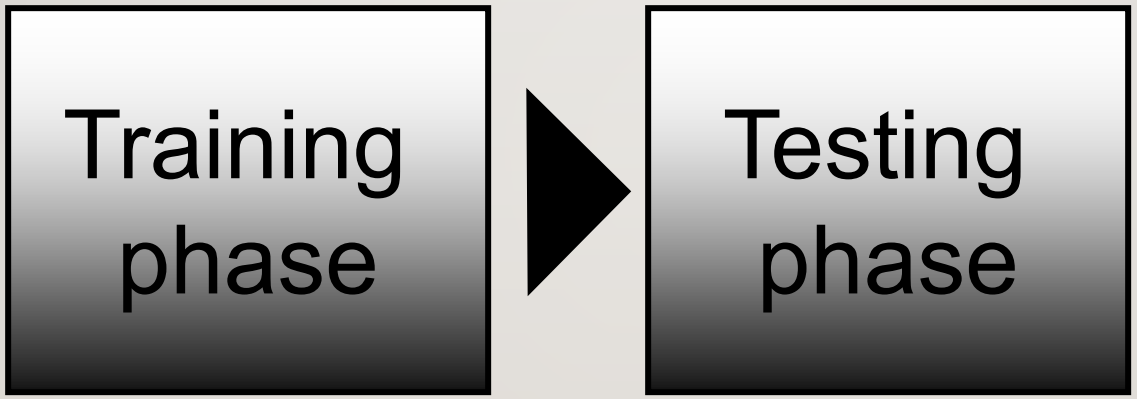
CLASSIFICATION

WHAT IS CLASSIFICATION?

A machine learning task that deals with identifying the class to which an instance belongs

A classifier performs classification

CLASSIFICATION LEARNING



Learning the classifier from the available data

'Training set'

(Labeled)

Testing how well the classifier performs

'Testing set'

AN EXAMPLE APPLICATION

- An emergency room in a hospital measures 17 variables (e.g., blood pressure, age, etc) of newly admitted patients.
- A decision is needed: whether to put a new patient in an intensive-care unit.
- Due to the high cost of ICU, those patients who may survive less than a month are given higher priority.
- Problem: to predict high-risk patients and discriminate them from low-risk patients.

ANOTHER APPLICATION

- A credit card company receives thousands of applications for new cards. Each application contains information about an applicant,
  - age
  - Marital status
  - annual salary
  - outstanding debts
  - credit rating
  - etc.
- Problem: to decide whether an application should approved, or to classify applications into two categories, approved and not approved.

THE DATA AND THE GOAL

- Data: A set of data records (also called examples, instances or cases) described by
  - *k* attributes: $A_1, A_2, \ldots A_k$.
  - a class: Each example is labelled with a pre-defined class.
- Goal: To learn a classification model from the data that can be used to predict the classes of new (future, or test) cases/instances.

AN EXAMPLE: DATA (LOAN APPLICATION)

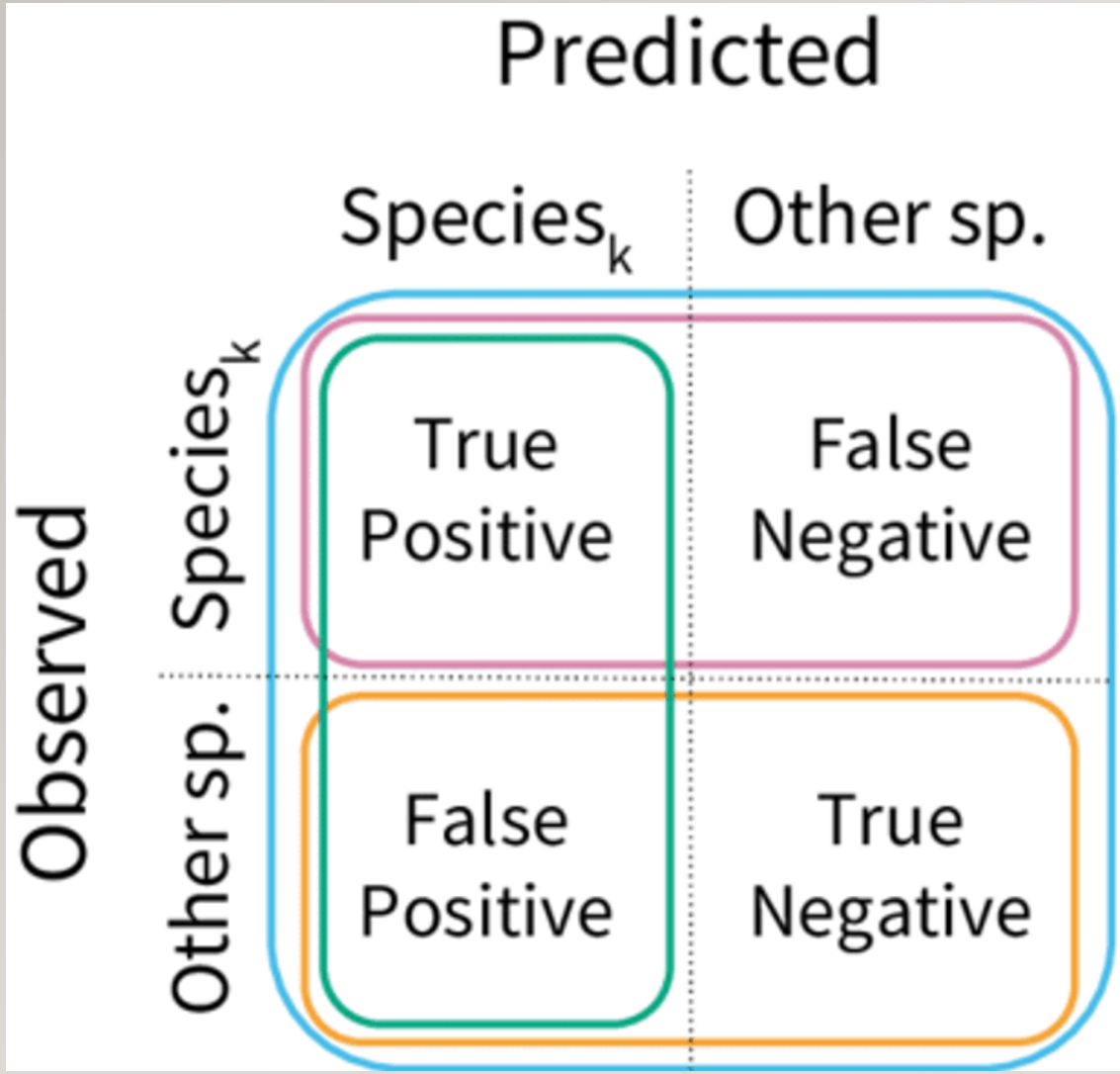
ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	young	false	false	fair	No
2	young	false	false	good	No
3	young	true	false	good	Yes
4	young	true	true	fair	Yes
5	young	false	false	fair	No
6	middle	false	false	fair	No
7	middle	false	false	good	No
8	middle	true	true	good	Yes
9	middle	false	true	excellent	Yes
10	middle	false	true	excellent	Yes
11	old	false	true	excellent	Yes
12	old	false	true	good	Yes
13	old	true	false	good	Yes
14	old	true	false	excellent	Yes
15	old	false	false	fair	No

Approved or not

AN EXAMPLE: THE LEARNING TASK

- Learn a classification model from the data
- Use the model to classify future loan applications into
  - Yes (approved) and
  - No (not approved)
- What is the class for following case/instance?

Age	Has_Job	Own_house	Credit-Rating	Class
young	false	false	good	?



Accuracy $= \frac{TP + TN}{TP + TN + FP + FN}$

Specificity $= \frac{TN}{TN + FP}$

Precision $= \frac{TP}{TP + FP}$

Recall $= \frac{TP}{TP + FN}$

CONCLUSION

- Logistic regression introduces us to a few things:
  - Alternate cost functions – depending on the scenario, we can switch between cost calculations.
  - Multiple class predictions – processes mostly the same, but the prediction processing is different.
- In sklearn, regularization (L2) is enabled by default.