# Convention, Accuracy metrics, Classification, Regression

Nipun Batra

December 26, 2023

IIT Gandhinagar

# Revision: What is Machine Learning

# Revision: What is Machine Learning

"Field of study that give computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

# Revision: What is Machine Learning

"Field of study that give computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

# Revision: What is Machine Learning

"Field of study that give computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

How would you program to recognise digits? Start with 4.

# Revision: What is Machine Learning

“Field of study that give computers the ability to learn without being explicitly programmed” - Arthur Samuel [1959]

How would you program to recognise digits? Start with 4.

Maybe 4 can be thought of as: “—” + “__” + “—” + another vertically down “—”

## Revision: What is Machine Learning

"Field of study that give computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

How would you program to recognise digits? Start with 4.

Maybe 4 can be thought of as: "—" + "__" + "—" + another vertically down "—"

The heights of each of the "—" need to be similar within tolerance

# Revision: What is Machine Learning

"Field of study that give computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

How would you program to recognise digits? Start with 4.

Maybe 4 can be thought of as: "—" + "__" + "—" + another vertically down "—"

The heights of each of the "—" need to be similar within tolerance

Each of the "—" can be slightly slanted. Similarly the horizontal line can be slanted.

# Revision: What is Machine Learning

"Field of study that give computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

How would you program to recognise digits? Start with 4.

Maybe 4 can be thought of as: "—" + "__" + "—" + another vertically down "—"

The heights of each of the "—" need to be similar within tolerance

Each of the "—" can be slightly slanted. Similarly the horizontal line can be slanted. There can be some cases of 4 where the first "—" is at 45 degrees

# Revision: What is Machine Learning

"Field of study that give computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

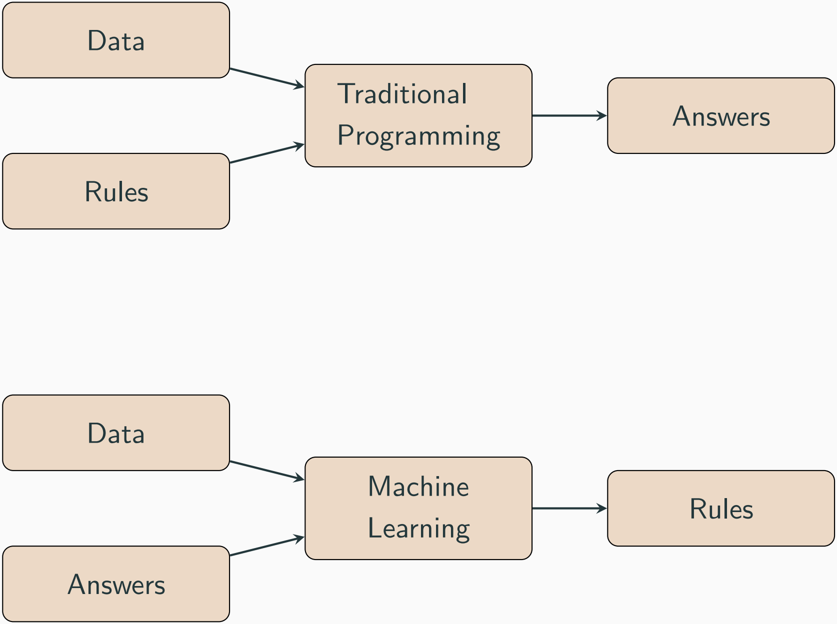
How would you program to recognise digits? Start with 4.

Maybe 4 can be thought of as: "—" + "__" + "—" + another vertically down "—"

The heights of each of the "—" need to be similar within tolerance

Each of the "—" can be slightly slanted. Similarly the horizontal line can be slanted. There can be some cases of 4 where the first "—" is at 45 degrees There can be some cases of 4 where the width of each stroke is different


Data
Rules
Traditional Programming
Answers


Data
Rules
Traditional Programming
Answers
Data
Answers
Machine Learning
Rules

# Revision: What is Machine Learning

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E." - Tom Mitchell

# First ML Task: Grocery store tomatoes quality prediction

Problem statement: You want to predict the quality/condition of a tomato given its visual features.

# Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

# Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

- Size

# Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

- Size
- Colour

# Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

- Size
- Colour
- Texture

# Dataset

Imagine you have some past data on quality of tomatoes.

| **Sample** | **Colour** | **Size** | **Texture** | **Condition** |
|---|---|---|---|---|
| 1 | Orange | Small | Smooth | Good |
| 2 | Red | Small | Rough | Good |
| 3 | Orange | Medium | Smooth | Bad |
| 4 | Yellow | Large | Smooth | Bad |

# Useful Features

Is the sample number a useful feature for predicting quality of a tomato?

# Useful Features

Is the sample number a useful feature for predicting quality of a tomato?

Answer: It depends! Maybe, all tomatoes received after a certain date are bad! Let us ignore that for now.

# Useful Features

Is the sample number a useful feature for predicting quality of a tomato?

Answer: It depends! Maybe, all tomatoes received after a certain date are bad! Let us ignore that for now.

Let us modify our data table for now.

| **Colour** | **Size** | **Texture** | **Condition** |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

# Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

# Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

The training set consists of two parts:

# Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

The training set consists of two parts:

1. Features, Attributes or Covariates

# Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

The training set consists of two parts:

1. Features, Attributes or Covariates
2. Output or Response Variable

# Training Set

| Colour | Size | Texture | Condition |
| --- | --- | --- | --- |
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

# Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

We call this matrix as $\mathcal{D}$, containing:

## Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathcal{R}^{\mathbf{N} \times \mathbf{P}}$) containing data of $N$ samples each of which is $P$ dimensional.

# Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathcal{R}^{\mathbf{N} \times \mathbf{P}}$) containing data of $N$ samples each of which is $P$ dimensional.
    - Thus, $\mathbf{X} = \{x_i^T\}_{i=1}^N$ where $x_i \in \mathcal{R}^P$

## Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathcal{R}^{\mathbf{N} \times \mathbf{P}}$) containing data of $N$ samples each of which is $P$ dimensional.
    - Thus, $\mathbf{X} = \{x_i^T\}_{i=1}^N$ where $x_i \in \mathcal{R}^P$
    - Example $x_1 = \begin{bmatrix} Orange \\ Small \\ Smooth \end{bmatrix}$

## Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathcal{R}^{\mathbf{N} \times \mathbf{P}}$) containing data of $N$ samples each of which is $P$ dimensional.
    - Thus, $\mathbf{X} = \{x_i^T\}_{i=1}^N$ where $x_i \in \mathcal{R}^P$
    - Example $x_1 = \begin{bmatrix} Orange \\ Small \\ Smooth \end{bmatrix}$
2. Output Vector ($y \in \mathcal{R}^N$) containing output variable for $N$ samples.

# Training Set

| Colour | Size | Texture | Condition |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathcal{R}^{\mathbf{N}\times\mathbf{P}}$) containing data of $N$ samples each of which is $P$ dimensional.
    - Thus, $\mathbf{X} = \{x_i^T\}_{i=1}^N$ where $x_i \in \mathcal{R}^P$
    - Example $x_1 = \begin{bmatrix} Orange \\ Small \\ Smooth \end{bmatrix}$
2. Output Vector ($y \in \mathcal{R}^N$) containing output variable for $N$ samples.
3. Thus, we can also write $\mathcal{D} = \{(x_i^T, y_i)\}_{i=1}^N$

# Prediction Task

Estimate condition for unseen tomatoes (#5, 6) based on data set.

| **Colour** | **Size** | **Texture** | **Condition** |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |
| Red | Large | Rough | ? |
| Orange | Large | Rough | ? |

# Testing Set

Testing set is similar to training set, but, does not contain labels for output variable.

| **Colour** | **Size** | **Texture** | **Condition** |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |
| Red | Large | Rough | ? |
| Orange | Large | Rough | ? |

## Prediction Task

We hope to:

# Prediction Task

We hope to:

1. Learn $f$: Condition $= f$(colour, size, texture)

## Prediction Task

We hope to:

1. Learn $f$: Condition $= f(\text{colour, size, texture})$
2. From Training Dataset

# Prediction Task

We hope to:

1. Learn $f$: Condition $= f$(colour, size, texture)
2. From Training Dataset
3. To Predict the condition for the Testing set

| **Colour** | **Size** | **Texture** | **Condition** |
|---|---|---|---|
| Orange | Small | Smooth | Good |
| Red | Small | Rough | Good |
| Orange | Medium | Smooth | Bad |
| Yellow | Large | Smooth | Bad |
| Red | Large | Rough | ? |
| Orange | Large | Rough | ? |

# Generalisation

- Q: Is predicting on test set enough to say our model generalises?

# Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!

# Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!
- Ideally - we want to predict "well" on all possible inputs. But, can we test that?

# Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!
- Ideally - we want to predict "well" on all possible inputs. But, can we test that?
- No! Since, the test set is only a sample from all possible inputs.

# Generalisation

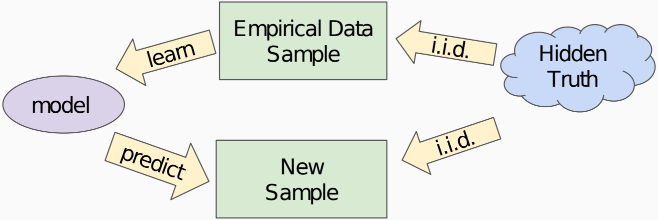



Image courtesy Google ML crash course

# Generalisation

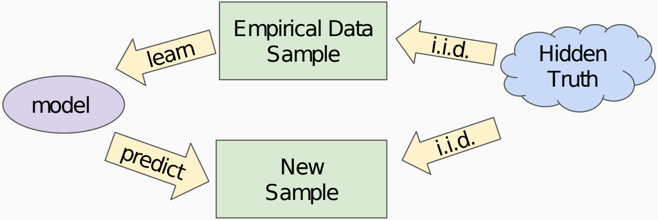



Image courtesy Google ML crash course

Both the training set and the test set are samples drawn from the hidden true distribution (also sometimes called population)

## Generalisation

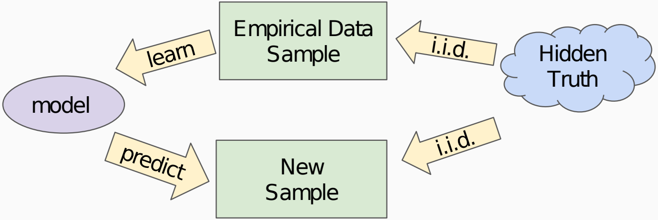



Image courtesy Google ML crash course

Both the training set and the test set are samples drawn from the hidden true distribution (also sometimes called population)

More discussion later once we study bias and variance

# Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

# Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

- # People (More people $\implies$ More Energy)

## Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

- # People (More people $\implies$ More Energy)
- Temperature (Higher Temp. $\implies$ Higher Energy)

# Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

- # People (More people $\implies$ More Energy)
- Temperature (Higher Temp. $\implies$ Higher Energy)

| # People | Temp (C) | Energy (kWh) |
|---|---|---|
| 4000 | 30 | 30 |
| 4200 | 30 | 32 |
| 4200 | 35 | 40 |
| 3000 | 20 | ? |
| 1000 | 45 | ? |

# Classification v/s Regression

- Classification

## Classification v/s Regression

- Classification
    - Output variable is discrete

# Classification v/s Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, \cdots C\}$

# Classification v/s Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, \cdots C\}$
  - Examples - Predicting:

## Classification v/s Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, \cdots C\}$
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)

## Classification v/s Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, \cdots C\}$
    - Examples - Predicting:
        - Will I get a loan? (Yes, No)
        - What is the quality of fruit? (Good, Bad)

# Classification v/s Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, \cdots C\}$
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression

# Classification v/s Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, \cdots C\}$
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous

# Classification v/s Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, \cdots C\}$
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathcal{R}$

## Classification v/s Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, \cdots C\}$
    - Examples - Predicting:
        - Will I get a loan? (Yes, No)
        - What is the quality of fruit? (Good, Bad)
- Regression
    - Output variable is continuous
    - i.e. $y_i \in \mathcal{R}$
    - Examples - Predicting:

# Classification v/s Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, \cdots C\}$
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathcal{R}$
  - Examples - Predicting:
    - How much energy will campus consume?

# Classification v/s Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, \cdots C\}$
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathcal{R}$
  - Examples - Predicting:
    - How much energy will campus consume?
    - How much rainfall will fall?

# Metrics for Classification

$$\begin{array}{c} \text{Prediction } (\hat{y}) \\ \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} \end{array} \qquad \begin{array}{c} \text{Ground Truth } (y) \\ \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Bad} \end{pmatrix} \end{array}$$

| | |
|---|---|
| Ground Truth: | From the actual training set |
| Prediction: | Made by the model |

# Accuracy

$$\text{Prediction } (\hat{y}) \qquad\qquad \text{Ground Truth } (y)$$

$$\begin{array}{c} \checkmark \\ \checkmark \\ \\ \\ \checkmark \end{array}\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} \qquad \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Bad} \end{pmatrix}$$

# Accuracy

$$
\begin{array}{cc}
 & \text{Prediction } (\hat{y}) \\
\begin{array}{c} \checkmark \\ \checkmark \\ \\ \\ \checkmark \end{array} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix}
\end{array}
\qquad
\begin{array}{c}
\text{Ground Truth } (y) \\
\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Bad} \end{pmatrix}
\end{array}
$$

$$
\begin{aligned}
\text{Accuracy} &= \frac{||y = \hat{y}||}{||y||} \\
&= \frac{3}{5} = 0.6
\end{aligned}
$$

# Types of Data: Imbalanced Classes

$$\begin{array}{r}\text{1 sample}\ \{\\ \text{100 samples}\ \left\{\begin{array}{c}\\ \\ \\ \\ \end{array}\right.\end{array}\begin{pmatrix}\text{Bad}\\ \text{Good}\\ \text{Good}\\ \dots\\ \text{Good}\end{pmatrix}\qquad\qquad \text{Imbalanced Classes}$$

## Types of Data: Imbalanced Classes

$$\begin{array}{r} \text{1 sample}\left\{ \vphantom{\text{Bad}} \right. \\ \text{100 samples}\left\{ \vphantom{\begin{array}{c}\text{Good}\\\text{Good}\\\dots\\\text{Good}\end{array}} \right. \end{array} \begin{pmatrix} \text{Bad} \\ \text{Good} \\ \text{Good} \\ \dots \\ \text{Good} \end{pmatrix} \qquad \text{Imbalanced Classes}$$

Cases for this:

- Cancer Screening
- Planet Detection

# Accuracy Metrics: Precision

$$
\begin{array}{ll}
 & \text{Prediction } (\hat{y}) \\
\rightarrow \checkmark & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} \\
\rightarrow \checkmark & \\
\rightarrow & \\
\rightarrow & 
\end{array}
\qquad
\begin{array}{c}
\text{Ground Truth } (y) \\
\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Good} \end{pmatrix}
\end{array}
$$

$$\text{Precision} = \frac{||y = \hat{y} = \text{Good}||}{||\hat{y} = \text{Good}||} = \frac{2}{4} = 0.5$$

"the fraction of relevant instances among the retrieved instances", i.e. "out of the number of times we predict Good, how many times is the condition actually Good"

# Accuracy Metrics: Precision

$$\begin{array}{c c} & \text{Prediction } (\hat{y}) \\ \begin{array}{l} \rightarrow \checkmark \\ \rightarrow \checkmark \\ \rightarrow \\ \rightarrow \\ \phantom{x} \end{array} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} \end{array} \qquad \begin{array}{c} \text{Ground Truth } (y) \\ \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Good} \end{pmatrix} \end{array}$$

$$\text{Precision} = \frac{||y = \hat{y} = \text{Good}||}{||\hat{y} = \text{Good}||} = \frac{2}{4} = 0.5$$

“the fraction of relevant instances among the retrieved instances”, i.e. “out of the number of times we predict Good, how many times is the condition actually Good”

## Accuracy Metrics: Recall

$$\begin{array}{cc} & \text{Prediction } (\hat{y}) \\ \rightarrow \checkmark & \\ \rightarrow \checkmark & \\ & \\ & \\ \rightarrow & \end{array} \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} \qquad \text{Ground Truth } (y) \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Good} \end{pmatrix}$$

$$\text{Recall} = \frac{||y = \hat{y} = \text{Good}||}{||y = \text{Good}||} = \frac{2}{3} = 0.67$$

"the fraction of the total amount of relevant instances that were actually retrieved"

# Types of Data: Imbalanced Classes

Given predictions of whether a tissue is cancerous or not ($n = 100$).

$$
\begin{array}{cc}
 & \text{Prediction } (\hat{y}) \\
\rightarrow & \left(\begin{array}{c} \text{Yes} \\ \text{No} \\ \text{No} \\ \dots \\ \text{No} \end{array}\right)
\end{array}
\qquad
\begin{array}{cc}
 & \text{Ground Truth } (y) \\
\begin{array}{c} \\ \\ \\ \\ \rightarrow \end{array} & \left(\begin{array}{c} \text{No} \\ \text{No} \\ \dots \\ \text{No} \\ \text{Yes} \end{array}\right) a
\end{array}
$$

# Types of Data: Imbalanced Classes

Given predictions of whether a tissue is cancerous or not ($n = 100$).

$$\begin{array}{c} \\ \rightarrow \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \text{Prediction } (\hat{y}) \\ \begin{pmatrix} \text{Yes} \\ \text{No} \\ \text{No} \\ \dots \\ \text{No} \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ \\ \\ \\ \\ \rightarrow \end{array} \begin{array}{c} \text{Ground Truth } (y) \\ \begin{pmatrix} \text{No} \\ \text{No} \\ \dots \\ \text{No} \\ \text{Yes} \end{pmatrix} a \end{array}$$

$$\text{Accuracy} = \frac{98}{100} = 0.98 \qquad \begin{aligned} \text{Recall} &= \frac{0}{1} = 0 \\ \text{Precision} &= \frac{0}{1} = 0 \end{aligned}$$

# Accuracy Metrics: Confusion Matrix

| | | Ground Truth | |
|---|---|---|---|
| | | Yes | No |
| Predicted | Yes | 0 | 1 |
| | No | 1 | 98 |

# Accuracy Metrics: Confusion Matrix

| | | Ground Truth | |
|---|---|---|---|
| | | Yes | No |
| Predicted | Yes | 0 | 1 |
| | No | 1 | 98 |

| | | Ground Truth | |
|---|---|---|---|
| | | Yes | No |
| Predicted | Yes | True Positive | False Positive |
| | No | False Negative | True Negative |

## Accuracy Metric: Confusion Matrix

| | | Ground Truth | |
|---|---|---|---|
| | | Yes | No |
| Predicted | Yes | True Positive | False Positive |
| | No | False Negative | True Negative |

$$\text{Precision} = \frac{T.P.}{T.P. + F.P.}$$

# Accuracy Metric: Confusion Matrix

| | | Ground Truth | |
|---|---|---|---|
| | | Yes | No |
| Predicted | Yes | True Positive | False Positive |
| | No | False Negative | True Negative |

$$\text{Precision} = \frac{T.P.}{T.P. + F.P.}$$

# Accuracy Metric: Confusion Matrix

| | | Ground Truth | |
|---|---|---|---|
| | | Yes | No |
| Predicted | Yes | True Positive | False Positive |
| | No | False Negative | True Negative |

$$\text{Recall} = \frac{T.P.}{T.P. + F.N.}$$

# Accuracy Metric: Confusion Matrix

| | | Ground Truth | |
|---|---|---|---|
| | | Yes | No |
| Predicted | Yes | True Positive | False Positive |
| | No | False Negative | True Negative |

$$\text{Recall} = \frac{T.P.}{T.P. + F.N.}$$

# Accuracy Metrics: F-Score

| | | Ground Truth | |
|---|---|---|---|
| | | Yes | No |
| Predicted | Yes | True Positive | False Positive |
| | No | False Negative | True Negative |

$$\mathrm{F-}\ \text{Score} = \frac{2\times\ \text{Precision}\ \times\ \text{Recall}}{\text{Precision} + \text{Recall}}$$

# Accuracy Metrics: Matthew's Correlation Coefficient

| | | Ground Truth | |
|---|---|---|---|
| | | Yes | No |
| Predicted | Yes | True Positive | False Positive |
| | No | False Negative | True Negative |

Matthew's correlation coefficient =

$$\frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

## Accuracy Metrics: Example

For the data given below, calculate:

$$\begin{array}{ll} & \text{G.T. Positive} \quad \text{G.T. Negative} \\ \begin{array}{l}\text{Pred Positive} \\ \text{Pred Negative}\end{array} & \begin{pmatrix} 90 & 4 \\ 1 & 1 \end{pmatrix} \end{array}$$

Precision $=$ ?

Recall $=$ ?

F-Score $=$ ?

Matthew's Coeff. $=$ ?

# Accuracy Metrics: Answer

For the same data

$$\begin{array}{r c}
 & \begin{array}{cc} \text{G.T. Positive} & \text{G.T. Negative} \end{array} \\
\begin{array}{r} \text{Pred Positive} \\ \text{Pred Positive} \end{array} & \begin{pmatrix} 90 & 4 \\ 1 & 1 \end{pmatrix}
\end{array}$$

Precision $= \frac{90}{94}$
Recall $= \frac{90}{91}$
F-Score $= 0.9524$
Matthew's Coeff. $= 0.14$

# Metrics for Regression MSE & MAE

$$\text{Prediction } (\hat{y}) \quad \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} \qquad \text{Ground Truth } (y) \quad \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix}$$

$$\text{Mean Squared Error (MSE)} = \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\text{MSE}}$$

# Accuracy Metrics: MAE & ME

$$\text{Prediction } (\hat{y}) \quad \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} \qquad \text{Ground Truth} \quad \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix}$$

$$\text{Mean Absolute Error (ME)} = \frac{\sum_{i=1}^{N} |\hat{y}_i - y_i|}{N}$$

$$\text{Mean Error} = \frac{\sum_{i=1}^{N} \hat{y}_i - y_i}{N}$$

# Accuracy Metrics: MAE & ME

$$\text{Prediction } (\hat{y}) \quad \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} \qquad \text{Ground Truth} \quad \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix}$$

$$\text{Mean Absolute Error (ME)} = \frac{\sum_{i=1}^{N} |\hat{y}_i - y_i|}{N}$$

$$\text{Mean Error} = \frac{\sum_{i=1}^{N} \hat{y}_i - y_i}{N}$$

Is there any downside with using mean error?

# Accuracy Metrics: MAE & ME

$$\text{Prediction } (\hat{y}) \quad \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} \qquad \text{Ground Truth} \quad \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix}$$

$$\text{Mean Absolute Error (ME)} = \frac{\sum_{i=1}^{N} |\hat{y}_i - y_i|}{N}$$

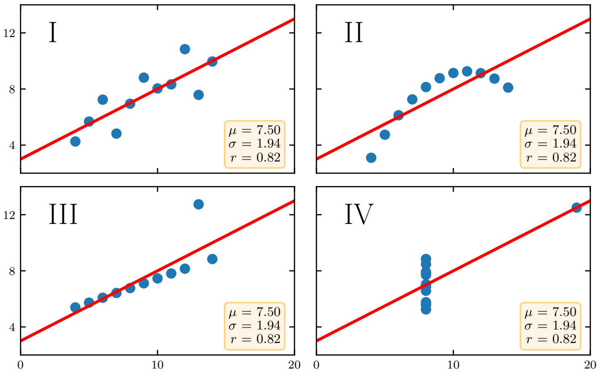
$$\text{Mean Error} = \frac{\sum_{i=1}^{N} \hat{y}_i - y_i}{N}$$

Is there any downside with using mean error?

Errors can get cancelled out

# The Importance of Plotting

## Notebook: anscombe.html




Anscombe's Quartet

# The Importance of Plotting

| Property | Value | Accross datasets |
|---|---|---|
| mean(X) | 9 | exact |
| mean(Y) | 7.5 | upto 3 decimal places |
| Linear regression line | $y = 3.00 + 0.500x$ | upto 2 decimal places |