UmdeMi

Summary of Statistics in ML

Overview

In this file, I will summarize some important topics on probability and statistics.

And know Why Statistics is given so much importance for excelling into Machine learning and artificial intelligence?

Goals

1. know the use of probability and statistics for computer vision.
2. Know some important terminology of probability.
3. Machine Learning uses Statistics for understanding
4. problems of statistical modeling in machine learning

Milestones

Probability and Statistics for Computer Vision

- Probability is using when making decisions.

- Probability is represented by % in real-life, but in math is represented using decimals.

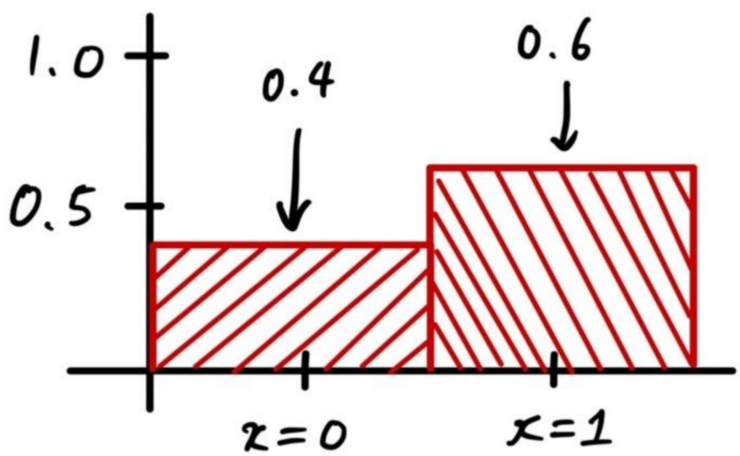
- ## Random variable RV :-

  - Is a variable representing the outcome of our interest.
  - Ex:-

    When you flip a coin, we donate to it by X which has two probability. P(X = 1) = 0.5 and P(X = 0) = 0.5

- ## Probability Density Function PDF

  - The sum of all the states(probabilities) is always 1



  - This is discrete PDF of dropping a card
  - The sum of all the areas are always 1.

- Joint probability

- Is a probability that calculates the likelihood of two or more events occurring together and at the same point in time.

- EX:-

You have 2 random variables x and y, x representing whether it rains or not, y representing whether you have an umbrella or not.

P(rain) --> $p(x = 1) = 0.6$	P(not rain) --> $p(x = 0) = 0.4$
P(have umbrella) --> $p(y = 1) = 0.3$	P(Don't have umbrella) --> $p(y = 0) = 0.7$

What's the probability that it rains and you have an umbrella?

$$P(x = 1 , y = 1) = p(x = 1) * p(y = 1)$$

$$= 0.6 * 0.3 = 0.18$$

- Marginalization

- Is a way to go from joint probability to the normal probability.
- We are looking for the **individual probabilities such as P(x=1) or P(y=0)**.

- we calculate that from these joint probabilities

$$Pr(x) = \int Pr(x, y)\, dy \qquad \text{(Continuous)}$$

$$Pr(x) = \sum_{y} Pr(x, y) \qquad \text{(Discrete)}$$

- Sum all the possible states for the RV you are not interested in.

- Ex:-

Suppose you want to get P(x=0), but you only have joint probabilities P (x=0, y=0) & P (x=0, y=1).

To get Pr(x=0)perform **marginalization** by doing the following:

$$\begin{aligned} Pr(x=0) &= \sum_{y=0}^{1} Pr(x=0, y) \\ &= Pr(x=0, y=0) + Pr(x=0, y=1) \\ &= 0.28 + 0.12 = 0.4 \end{aligned}$$

- Conditional Probability

- is a measure of the probability of an event given that another event has already occurred.

Given y, probability of x

Joint probability of x & y

$$Pr(x \mid y) = \frac{Pr(x, y)}{Pr(y)}$$

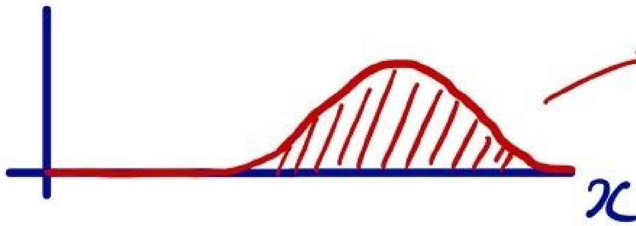
Probability of y
(Normalization Term)

- We calculate it by dividing the joint probability by the probability of the state where you are basing your decision on.
- **What is meant by normalization term:-**

All the possible joint probabilities in this area have to sum to 1.

That's why taking a slice won't give you 1 since you are ignoring all, the other possible joint probability cases.

Therefore, to make this slice a PDF, we need to normalize the area so that this slice area becomes 1.



What is this area?

$$\int Pr(x, y = y^*)\,dx$$

$$Pr(x \mid y = y^*) = \frac{Pr(x, y = y^*)}{\int Pr(x, y = y^*)\,dx}$$

← Normalization Term (By dividing with it's own area, normalizing to 1.)

$$= \frac{Pr(x, y = y^*)}{Pr(y = y^*)}$$

- **Bayes' rule**

Bayes' rule Formula:-

Bayes' rule

$$\boxed{Pr(y \mid x) = \frac{Pr(x \mid y)\, Pr(y)}{Pr(x)}}$$

$$\boxed{Pr(y|x) = \frac{Pr(x|y)\,Pr(y)}{Pr(x)}}$$

$$= \frac{Pr(x|y)\,Pr(y)}{\int Pr(x,y)\,dy}$$

$\because$ Marginalization

$Pr(x) = \int Pr(x,y)\,dy$

$$= \frac{Pr(x|y)\,Pr(y)}{\int Pr(x|y)\,Pr(y)\,dy}$$

$\because$ Conditional Probability

$Pr(x,y) = Pr(x|y)\,Pr(y)$

Calculate the posterior using 3 terms:

- **Likelihood**

  propensity for observing a certain value of x given a certain value of y

- **Prior**

  what we know about y before observing x

- **Evidence**

  a constant to ensure that the left hand side is a valid distribution (normalization term)

$$\underbrace{Pr(y|x)}_{\text{Posterior}} = \frac{\overbrace{Pr(x|y)}^{\text{Likelihood}}\,\overbrace{Pr(y)}^{\text{Prior}}}{\underbrace{\int Pr(x|y)\,Pr(y)\,dy}_{\text{Evidence}}}$$

From Conditional Probability

$$\begin{cases} Pr(x|y) = \dfrac{Pr(x,y)}{Pr(y)} \\[2ex] Pr(y|x) = \dfrac{Pr(x,y)}{Pr(x)} \end{cases}$$

$$\begin{cases} Pr(x,y) = Pr(x|y)\,Pr(y) \cdots ① \\ Pr(x,y) = Pr(y|x)\,Pr(x) \cdots ② \end{cases}$$

$$② = ①$$

$$Pr(y|x)\,Pr(x) = Pr(x|y)\,Pr(y)$$

$$\therefore \boxed{Pr(y|x) = \dfrac{Pr(x|y)\,Pr(y)}{Pr(x)}}$$

How is Machine Learning Different from Statistics and Why it Matters

- **Statistics** a subfield of Mathematics. it’s a formalization of relationships between variables in the data in the form of mathematical equations.
- There are Common estimator include: P-value, Standard deviation, Confidence interval, and Unbiased estimator
- **Machine Learning (ML)** is a subfield of computer science and AI that deals with building systems that can learn from data and observations.
- **Differences between Statistics and Machine Learning.**

	Statistics	Machine Learning
Approach	Data Generating Process	Algorithmic Model
Driver	Math, Theory	Fitting Data
Focus	Hypothesis Testing, Interpretability	Predictive Accuracy
Data Size	Any Reasonable Set	Big Data
Dimensions	Used Mostly for Low Dimensions	High Dimensional Data
Inference	Parameter Estimation, Predictions, Estimating Error Bars	Prediction
Model Choice	Parameter Significance, In-sample Goodness of Fit	Cross-validation of Predictive Accuracy on Partitions of Data
Popular Tools	R	Python
Interpretability	High	Low

- **Feature Engineering**
  - Large number of inputs
  - Unstructured data needs feature engineering as a pre-processing step before training.
  - AutoML can generate a large number of complex features to test many transformations of the data.
- **Hyperparameters**

  Must be defined before the training process can begin

  It can be:-

  - The depth of trees in a random-forest
  - The number of layers in a deep neural
  - The Learning rate
  - Number of epochs
  - Batch Size

- **Statistics in the ML Workflow**

  - **Statistics** is key to data preparation and sound validation and often used as part of the modeling process.
  - **Data Exploration** is the first step performed on data, is informed by statistics.
  - **Validation** The accepted approach is to check p-value of a sample to ensure its over 5% confidence level it passes the significance test.

★ **Lost in Translation** Many times Statistical modeling and ML use very similar approaches and therefore overlap with each other as logistic regression.

Statistics	Machine Learning
Data Point	Instance
Covariate	Feature
Parameters	Weights
Estimation / Fitting	Learning
Regression / Classification	Supervised Learning
Clustering / Density Estimation	Unsupervised Learning
Response	Label
Test set performance	Generalization