

3.4 PROBABILISTIC GENERATIVE MODELS

- Generative models are a class of statistical models that generate new data instances. These models are used in unsupervised machine learning to perform tasks such as probability and likelihood estimation, modelling data points, and distinguishing between classes using these probabilities,
- Generative models rely on the Bayes theorem to find the joint probability. Generative models describe how data is generated using probabilistic models. They predict $P(y/x)$, the probability of y given x , calculating the $P(x,y)$, the probability of x and y .

Naive Bayes

Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong independence assumptions between the features. It is highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem.

A Naive Bayes Classifier is a program which predicts a class value given a set of attributes.

For each known class value,

1. Calculate probabilities for each attribute, conditional on the class value.
2. Use the product rule to obtain a joint conditional probability for the attributes.
3. Use Bayes rule to derive conditional probabilities for the class variable.
 - Once this has been done for all class values, output the class with the highest probability.
 - Naive bayes simplifies the calculation of probabilities by assuming that the probability of each attribute belonging to a given class value is independent of all other attributes. This is a strong assumption but results in a fast and effective method.
 - The probability of a class value given a value of an attribute is called the conditional probability. By multiplying the conditional probabilities together for each attribute for a given class value, we have a probability of a data instance belonging to that class.

Conditional Probability

Let A and B be two events such that $P(A) > 0$. We denote $P(B/A)$ the probability of B given that A has occurred. Since A is known to have occurred, it becomes the new sample space replacing the original S. From this, the definition is,

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

OR

$$P(A \cap B) = P(A) \cdot P(B/A)$$

- The notation $P(B | A)$ is read "the probability of event B given event A". It is the probability of an event B given the occurrence of the event A.
- We say that, the probability that both A and B occur is equal to the probability that A occurs times the probability that B occurs given that A has occurred. We call $P(B/A)$ the conditional probability of B given A, i.e., the probability that B will occur given that A has occurred.
- Similarly, the conditional probability of an event A given B by.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

- The probability $P(A/B)$ simply reflects the fact that the probability of an event A may depend on a second event B. If A and B are mutually exclusive $A \cap B = \emptyset$ and $P(A/B) = 0$. Another way to look at the conditional probability formula is:

$$P(\text{Second/First}) = \frac{P(\text{First choice and second choice})}{P(\text{First choice})}$$

- Conditional probability is a defined quantity and cannot be proven.
- The key to solving conditional probability problems is to:

1. Define the events
2. Express the given information and question in probability notation.

3. Apply the formula.

Joint Probability

- A joint probability is a probability that measures the likelihood that two or more events will happen concurrently.
- If there are two independent events A and B, the probability that A and B will occur is found by multiplying the two probabilities. Thus for two events A and B, the special rule of multiplication shown symbolically is:

$$P(\mathbf{A \text{ and } B}) = P(\mathbf{A}) P(\mathbf{B}).$$

- The general rule of multiplication is used to find the joint probability that two events will occur. Symbolically, the general rule of multiplication is,

$$P(\mathbf{A \text{ and } B}) = P(\mathbf{A}) P(\mathbf{B/A}).$$

. The probability $P(A \cap B)$ is called the joint probability for two events A and B which intersect in the sample space. Venn diagram will readily shows that

$$P(\mathbf{A \cap B}) = P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A \cup B})$$

Equivalently

$$P(\mathbf{A \cap B}) = P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A \cap B}) \leq P(\mathbf{A}) + P(\mathbf{B})$$

The probability of the union of two events never exceeds the sum of the event probabilities.

A tree diagram is very useful for portraying conditional and joint probabilities. A tree diagram portrays outcomes that are mutually exclusive.

Bayes Theorem

- Bayes' theorem is a method to revise the probability of an event given additional information. Bayes's theorem calculates a conditional probability called a posterior or revised probability.

- Bayes' theorem is a result in probability theory that relates conditional probabilities. If A and B denote two events, $P(A|B)$ denotes the conditional probability of A occurring, given that B occurs. The two conditional probabilities $P(A|B)$ and $P(B|A)$ are in general different.
- Bayes theorem gives a relation between $P(A|B)$ and $P(B|A)$. An important application of Bayes' theorem is that it gives a rule how to update or revise the strengths of evidence-based beliefs in light of new evidence a posteriori.
- A **prior probability** is an initial probability value originally obtained before any additional information is obtained.
- A **posterior probability** is a probability value that has been revised by using additional information that is later obtained.

Suppose that $B_1, B_2, B_3, \dots, B_n$ partition the outcomes of an experiment and that A is another event. For any number, k, with $1 \leq k \leq n$, we have the formula:

$$P(B_k/A) = \frac{P(A/B_k) \cdot P(B_k)}{\sum_{i=1}^n P(A/B_i) \cdot P(B_i)}$$

Difference between Generative models and Discriminative models

Criteria	Generative Models	Discriminative Models
Application (Use cases)	Generate new data instances. e.g.- Predict next word in a sequence.	Discriminate between different kinds of data instances e.g.- Text Classification.
Model Performance	<ul style="list-style-type: none"> • Measured by likelihood. • Outliers affect the distribution significantly and results in bad model accuracy. • With correct distribution the generative models are more accurate. 	<ul style="list-style-type: none"> • Measured by misclassification rate. • These models are robust against outliers and poor modeling.
Training time	Need less data to train as these models make stronger assumptions of conditional independence	Discriminative models take more time to train on substantial amount of training samples.
Training Data	Generative models can work with missing data and generalize well	Discriminative Models need sufficient training data for better generalization performance.
Intuitive	More elegant by having explanatory power	Relationships between variables are not explicit and visualizable (Black boxes)