1 Basic Probability

1.1 Joint probability

The probability of two (or more) events is called the joint probability. The joint probability of two or more random variables is referred to as the joint probability distribution.

X and Y as random varibles with events A and B. Joint probability is given by **product rule** of probability or the **chain rule** of probability :

$$P(A \text{ and } B) = P(A \text{ given } B) * P(B)$$
$$\text{and}$$
$$P(B \text{ and } A) = P(B \text{ given } A) * P(A)$$

with symmetry,

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

$P(A \text{ given } B)$ is **conditional probability**

1.2 Marginal probability

the probability of $X = A$ for all outcomes of Y. It is simply union over all the probabilities of all events for the second variable for a given fixed event for the first variable.

$$P(X = A) = \sum_i P(X = A, Y = y_i) \; \forall y$$

1.3 Conditional probability

$P(A \text{ given } B)$ also written as $P(A|B)$. From Joint probability,

$$P(A \text{ and } B) = P(A \text{ given } B) * P(B)$$
$$=>$$
$$P(A \text{ given } B) = P(A \text{ and } B)/P(B)$$

1.4 Independence

If one variable is not dependent on a second variable, this is called independence or statistical independence.

- Joint probability: $P(A \text{ and } B) = P(A) * P(B)$
- Marginal probability: $P(A)$
- Conditional probability: $P(A \text{ given } B) = P(A)$

1.5 Exclusivity or Disjoint events

- Joint probability: $P(A \text{ and } B) = 0$
- $P(A \text{ or } B) = P(A) + P(B) - P(A, B)$ gets reduced to $P(A \text{ or } B) = P(A) + P(B)$

2 Bayes Theorem

2.1 Definition

From definition of Joint probability:

- based on conditional probability
- it is symmetric.

One could collect conditional probability terms and define a relation between the two condiitonal probabilities.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

with,

$$\text{posterior} = \frac{\text{likelihood x prior}}{\text{evidence}}$$

way of calculating a conditional probability without the joint probability. To get $P(B)$ indirectly, it can be reformulated as:

$$P(A|B) = \frac{P(B|A)P(A)}{(P(B|A)P(A) + P(B|\text{not}A)P(\text{notA}))} \tag{1}$$

2.2 Example

2.2.1 Diagnosis

- Cancer : Yes/No
- Diagnosis: Yes/No

Problem : Given
a) The test has a true positive rate or sensitivity of 85%.
b) test has true negative rate or specificity 95%
c) Prevalance of cancer is 0.002%
If a randomly selected patient has the test and it comes back positive, what is the probability that the patient has cancer?

Solution : From Bayes Theorem: We know,

$$P(Cancer = True|Test = Positive) = \frac{P(Test = Positive|Cancer = True) * P(Cancer = True)}{P(Test = Positive)}$$

$$P(Cancer = True|Test = Positive) = \frac{0.85 * 0.0002}{P(Test = Positive)}$$

To find $P(Test = Positive)$, from above section:

$$\begin{aligned} P(B) &= P(B|A) * P(A) + P(B|notA) * P(notA) \\ P(Test = Positive) &= P(Test = Positive|Cancer = True) * P(Cancer = True) \\ &\quad + P(Test = Positive|Cancer = False) * P(Cancer = False) \end{aligned}$$

where,

$$\begin{aligned} P(Cancer = False) &= 1 - P(Cancer = True) \\ &= 1 - 0.0002 \\ P(Test = Positive|Cancer = False) &= 1–P(Test = Negative|Cancer = False) \\ &= 1 - \text{specificity} \\ &= 1 - 0.95 \end{aligned}$$

We now have enough information to estimate the probability of a randomly selected person having cancer if they get a positive test result.

$$\begin{aligned} P(Cancer = True|Test = Positive) &= \frac{P(Test = Positive|Cancer = True) * P(Cancer = True)}{P(Test = Positive)} \\ &= \frac{\text{sensitivity * prevalance}}{(\text{sensitivity * prevalance}) + (1 - \text{specificity}) * (1 - \text{prevalance})} \\ &= \frac{0.85 * 0.002}{(0.85 * 0.0002) + (0.05 * 0.9998)} \\ &= 0.003389 \end{aligned}$$

if the patient is informed they 1have cancer with this test, then there is only 0.33% chance that they have cancer.

3 Bayes Classifier

4 Naive Bayes

It is :

- supervised learning using Bazes theoem
- with the "naive" assumption of conditional independence between every pair of features given class label.

Given class label y and dependent feature vector x_1 through x_n

$$P(y|x_1, ..., x_n) = \frac{P(x_1, ..., x_n|y)P(y)}{P(x_1, ..., x_n)}$$

Using the naive conditional independence assumption and removing constant denominator for a given data set, it is simplified to

$$\begin{aligned} P(y|x_1, ..., x_n) &= \frac{P(x_1|y)......P(x_n|y)P(y)}{P(x_1, ..., x_n)} \\ &\propto \prod_{i=1}^{n} P(x_i|y) \end{aligned}$$

and we can use Maximum A Posteriori (MAP) estimation to estimate $P(y)$ and $P(x_i|y)$.
This can be expressed as

$$y = \underset{y}{\operatorname{argmax}} \prod_{i=1}^{n} P(x_i|y)$$