Machine Learning

Naïve Bayes Classifier

Dr. Pratyay Kuila

Dept. of Computer Science & Engineering
NIT Sikkim, Ravangla-737139

Bayesian Learning: Introduction

- $P(x)$ denotes the probability of x. $P(x)$ is often called the *prior probability* of x.
- $P(x|y)$ denotes the probability of x given y.

- In machine learning, let D be the data instance and y be the class. We will write $P(D|y)$ to denote the probability of observing data D given some world in which class y holds.
- We are interested in the probability $P(y|D)$ that y holds given the observed data instance D.

- $P(y|D)$ is called the *posterior probability* of y, because it reflects our confidence that y holds after we have seen the training data D. Notice the posterior probability $P(y|D)$ reflects the influence of the training data D, in contrast to the prior probability $P(y)$, which is independent of D.

Probability and Bayes Theorem

- If X and Y are random variables, then $P(X = x, Y = y)$ denotes their *joint probability.*

- It refers to the probability that variable X will take on the value x **AND** variable Y will take on the value y.

- The random variables are *independent* of each other if $P(X,Y) = P(X) \times P(Y)$.

- If two random variables are independent, it means that the value for one variable has no impact on the value for the other.

- What if the random variables are dependent?

Probability and Bayes Theorem

- The conditional probability $P(Y = y|X = x)$ refers to the probability that the variable Y will take on the value y, given that the variable X is observed to have the value x.
- The joint and conditional probabilities for X and Y are related in the following way:

$$P(X,Y) = P(X|Y) \times P(Y) = P(Y|X) \times P(X)$$

- Conditional probability is useful concept for understanding the *dependencies among random variables.*
- The conditional probability for variable X given Y is defined as,

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

- If X and Y are independent, then $P(Y|X) = P(Y)$.

Probability and Bayes Theorem

- ***Bayes theorem*:** The conditional probabilities $P(Y|X)$ and $P(X|Y)$ can be expressed in terms of one another formula known as the *Bayes theorem*.

$$P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)}$$

- If $\{x_1, x_2, \ldots, x_k\}$ is the set of mutually exclusive and exhaustive outcomes of a random variable X, such that $\sum_{i=1}^{k} P(x_i) = 1$, then the denominator of the above equation can be expressed as follows:

$$P(X) = \sum_{i=1}^{k} P(X, x_i) = \sum_{i=1}^{k} P(X|x_i) \times P(x_i)$$

This is known as the *law of total probability*.

Bayes Theorem: Example

Example 1: A doctor knows that meningitis causes stiff neck 50% of the time. Prior probability of any patient having meningitis is 1/50,000. Prior probability of any patient having stiff neck is 1/20. If a patient has stiff neck, what is the probability he/she has meningitis?

Bayes Theorem: Example

Example 1: A doctor knows that meningitis causes stiff neck 50% of the time. Prior probability of any patient having meningitis is 1/50,000. Prior probability of any patient having stiff neck is 1/20. If a patient has stiff neck, what is the probability he/she has meningitis?

- $P(M)$ denotes the probability of having meningitis (M).
- $P(S)$ denotes the probability of having stiff neck (S).
- **We have to find out $P(M|S)$**, the conditional probability of meningitis (M) given the patient has stiff neck (S).
- ***Given***: $P(S \mid M) = 0.5$, $P(M) = 1/50{,}000$ and $P(S) = 1/20$.

Bayes Theorem: Example

Example 1: A doctor knows that meningitis causes stiff neck 50% of the time. Prior probability of any patient having meningitis is 1/50,000. Prior probability of any patient having stiff neck is 1/20. If a patient has stiff neck, what is the probability he/she has meningitis?

- *P*(*M*) denotes the probability of having meningitis (*M*).
- *P*(*S*) denotes the probability of having stiff neck (*S*).
- **We have to find out *P*(*M*|*S*)**, the conditional probability of meningitis (*M*) given the patient has stiff neck (*S*).
- ***Given***: *P*(*S* | *M*) = 0.5, *P*(*M*) = 1/50,000 and *P*(*S*) = 1/20.
- Therefore, $P(M|S) = \frac{P(S|M) \times P(M)}{P(S)} = \frac{0.5 \times \frac{1}{50000}}{\frac{1}{20}} = 0.0002$

Bayes Theorem: Example

Example 2: Consider a football game between two teams A and B? Suppose team A wins 65% of the times and team B wins the remaining matches. Among the game won by team A, only 30% of them came from playing on team B's football field. On the other hand, 75% of the victories for team B are obtained while playing at home. If team B is to host the next match between the two teams, which team will most likely win the match?

Bayes Theorem: Example

Example 2: Consider a football game between two teams A and B? Suppose team A wins 65% of the times and team B wins the remaining matches. Among the game won by team A, only 30% of them came from playing on team B's football field. On the other hand, 75% of the victories for team B are obtained while playing at home. If team B is to host the next match between the two teams, which team will most likely win the match?

- Let X be the random variable represents the team hosting the match and Y be the random variable that represents the winner of the match.
- *Given*:
  - Probability team A wins is $P(Y=A) = 0.65$.
  - Probability team B wins is $P(Y=B) = 1- P(Y=A) = 0.35$.
  - Probability team B hosted the match it won is $P(X=B|Y=B) = 0.75$.
  - Probability team B hosted the match won by team A is: $P(X=B|Y=A) = 0.3$.
- We have to compute $P(Y=B|X=B)$.

Bayes Theorem: Example

Example 2 (Cont…):

- $P(Y=B|X=B)$ is the conditional probability that team B wins the next match it will be hosting, and compares it against $P(Y=A|X=B)$.
- Using the Bayes theorem, we obtain

$$P(Y=B|X=B) = \frac{P(X=B|Y=B)\times P(Y=B)}{P(X=B)} = \frac{P(X=B|Y=B)\times P(Y=B)}{P(X=B,Y=B)+P(X=B,Y=A)}$$

$$= \frac{P(X=B|Y=B).P(Y=B)}{P(X=B|Y=B).P(Y=B)+P(X=B|Y=A).P(Y=A)}$$

$$= \frac{0.75\times 0.35}{0.75\times 0.35 + 0.3\times 0.65} = 0.5738$$

- The law of total probability is applied in the second line.
- Furthermore, $P(Y=A|X=B) = 1 - P(Y=B|X=B) = 1 - 0.5738 = 0.4262$.
- Since, $P(Y=B|X=B) > P(Y=A|X=B)$, team B has a better chance than team A of winning the next match.

Bayes Theorem for Classification

- Consider each attribute and class label as random variable.
- Given a record with attributes $X = (x_1, x_2, \ldots, x_d)$
  - Goal is to predict class $Y=y$
  - Specifically, we want to find the value of Y that maximizes $P(Y=y \mid x_1, x_2, \ldots, x_d)$

- Can we estimate $P(Y=y \mid x_1, x_2, \ldots, x_d)$ directly from data?

Approach:

- Compute the posterior probability $P(Y=y \mid x_1, x_2, \ldots, x_d)$ for all values of Y using the Bayes theorem.
- Choose value of Y that maximizes $P(Y=y \mid x_1, x_2, \ldots, x_d)$.
- Equivalent to choosing value of Y that maximizes $P(x_1, x_2, \ldots, x_d \mid Y=y) \times P(Y=y)$

- How to estimate $P(x_1, x_2, \ldots, x_d \mid Y=y)$?

Naïve Bayes Classifier

- A Naïve Bayes classifier estimates the class-conditional probability by *assuming that the attributes are conditionally independent*, given the class label *y*. The conditional independence assumption can be formally stated as follows:

$$P(X|Y=y) = \prod_{i=1}^{d} P(x_i|Y=y)$$

where each attribute set $X=\{x_1, x_2,\ldots, x_d\}$ consists of *d* attributes.

- Naïve Bayes classifier assumes conditional independence between all x_i.

Naïve Bayes Classifier

- With the *conditional independence assumption*, instead of computing the class-conditional probability for every combination of *X*, only to estimate the conditional probability of each x_i, given *Y*.
- The Naïve Bayes classifier computes the posterior probability for each class *Y*=*y*.

$$P(Y=y|X) = \frac{P(Y=y)P(X|Y=y)}{P(X)} = \frac{P(Y=y)\prod_{i=1}^{d} P(x_i|Y=y)}{P(X)}$$

- Since *P*(*X*) is fixed for every *Y*=*y*, it is sufficient to choose the class that maximizes the numerator term, $P(Y=y) \times \prod_{i=1}^{d} P(x_i|Y=y)$

Assume independence among attributes x_i when class is given:

- $P(x_1, x_2, \ldots, x_n \mid Y=y) = P(x_1 \mid y)\, P(x_2 \mid y) \ldots P(x_n \mid y)$
- Can estimate $P(x_i \mid y)$ for all x_i and *y*.
- New point is classified to *y* if $P(Y=y)\prod_{i=1}^{d} P(x_i|Y=y)$ is maximal.

Naïve Bayes Classifier: Estimate Probabilities

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Class: $P(Y) = N_c / N$
  - e.g., $P(Y = No) = 7/10$ and $P(Y = Yes) = 3/10$
- *For categorical/ discrete attributes:* for categorical attribute A_i, the conditional probability $P(A_i = a_{ij} | C = c_k)$ is the fraction of training instances in class c_k that take on the attribute value a_{ij}.
  - $P(A_i = a_{ij} | C = c_k) = N_i / N_k$
  - where N_i is number of instances having attribute value $A_i = a_{ij}$ and belongs to class c_k. N_k be the total number of training instances for class c_k.

- *Examples*: four out of seven people belongs to *No* class are married, i.e., $P(Status = Married \mid No) = 4/7$. Similarly, the conditional probability of *Home Owner* =*Yes* belongs to *Yes* class is $P(Home = Yes | Yes) = 0$.

Naïve Bayes Classifier: Estimate Probabilities

- For continuous attributes:
  - Discretize the range into bins
    - One ordinal attribute per bin
    - Violates independence assumption
  - Two-way split: ($A < v$) or ($A > v$)
    - Choose only one of the two splits as new attribute
  - Probability density estimation:
    - Assume attribute follows a *normal distribution*.
    - Use data to estimate parameters of distribution (e.g., *mean* and *standard deviation*)
    - Once probability distribution is known, can use it to estimate the conditional probability $P(A_i|c)$

Naïve Bayes Classifier: Estimate Probabilities

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- *Normal distribution*:

$$P(x_i = a_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(a_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

  - One for each (x_i, y_j) pair

- For (*Income*, *Class* = *No*):
  - If *Class* = *No*
    - Sample mean $\mu = 110$
    - Sample variance $\sigma^2 = 2975$

$$P(Income = 120 | No) = \frac{1}{\sqrt{2\pi \times 2975}} e^{-\frac{(120-110)^2}{2\times 2975}} = 0.0072$$

Naïve Bayes Classifier: Example

Example 3: *Given a Test Record*: X = (Home = No, Married, Income = 120K)

P(Home Owner=Yes|No) = 3/7
P(Home Owner=No|No) = 4/7
P(Home Owner=Yes|Yes) = 0
P(Home Owner=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No) = 1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/3
P(Marital Status=Divorced|Yes) = 1/3
P(Marital Status=Married|Yes) = 0

For Annual Income:
If class=No:
  sample mean=110
  sample variance=2975
If class=Yes:
  sample mean=90
  sample variance=25

- $P(X|\text{Class=No})$

  $= P(\text{Home=No}|\text{Class=No}) \times P(\text{Married}|\text{Class=No}) \times P(\text{Income=120K}|\text{Class=No})$

  $= 4/7 \times 4/7 \times 0.0072 = 0.0024$

- $P(X|\text{Class=Yes})$

  $= P(\text{Home =No}|\text{Class=Yes}) \times P(\text{Married}|\text{Class=Yes}) \times P(\text{Income=120K}|\text{Class=Yes})$

  $= 1 \times 0 \times 1.2 \times 10^{-9} = 0$

Naïve Bayes Classifier: Example

Example 3: *Given a Test Record*: *X* = (Home = No, Married, Income = 120K)

- Therefore, $P(X|\text{No}) = 0.0024$ and $P(X|\text{Yes}) = 0$.
- The posterior probability for class *No* is

$$P(No|X) = \frac{P(X|No) \times P(No)}{P(X)} = \frac{0.0024 \times \frac{7}{10}}{P(X)} = 0.0016\alpha$$

where $\alpha = 1/P(X)$ is a constant term.

- Similarly, the posterior probability for class *Yes* is

$$P(Yes|X) = \frac{P(X|Yes) \times P(Yes)}{P(X)} = \frac{0 \times \frac{3}{10}}{P(X)} = 0$$

- Since, $P(No|X) > P(Yes|X)$, Class = No.

Naïve Bayes Classifier: Example

Example 4: Let us apply the Naïve Bayes classifier to the *PlayTennis* problem (refer to Table 1 of Decision Tree ppts). The Table provides a set of 14 training examples of the target concept *PlayTennis*, where each day is described by the attributes *Outlook*, *Temperature*, *Humidity*, and *Wind*.

- Here we use the Naïve Bayes classifier and the training data from this table to classify the following instance:

  (*Outlook* = *sunny*, *Temperature* = *cool*, *Humidity* = *high*, *Wind* = *strong*).

Naïve Bayes Classifier: Example

Example 4: Let us apply the Naïve Bayes classifier to the *PlayTennis* problem (refer to Table 1 of Decision Tree ppts). The Table provides a set of 14 training examples of the target concept *PlayTennis*, where each day is described by the attributes *Outlook*, *Temperature*, *Humidity*, and *Wind*.

- Here we use the Naïve Bayes classifier and the training data from this table to classify the following instance:

  (*Outlook* = *sunny*, *Temperature* = *cool*, *Humidity* = *high*, *Wind* = *strong*).

- Our task is to predict the target value (*yes* or *no*) of the target concept *PlayTennis* for this new instance.
- The target class value V_{NB} is given by

$$V_{NB} = argmax_{(y \in \{yes,no\})} P(Y = y) \prod_{i=1}^{d} P(x_i | Y = y)$$

Naïve Bayes Classifier: Example

$$V_{NB} = argmax_{(y \in \{yes,no\})} P(Y=y) \prod_{i=1}^{d} P(x_i|Y=y)$$

$$= argmax_{(y \in \{yes,no\})} P(Y=y).P(Outlook=sunny|Y=y)$$
$$.P(Temperature=cool|Y=y)$$
$$.P(Humidity=high|Y=y)$$
$$.P(Wind=strong|Y=y)$$

Table 1: Training Data for *PlayTennis*.

Day	***Outlook***	***Temperature***	***Humidity***	***Wind***	***PlayTennis***
D_1	*Sunny*	*Hot*	*High*	*Weak*	*No*
D_2	*Sunny*	*Hot*	*High*	*Strong*	*No*
D_3	*Overcast*	*Hot*	*High*	*Weak*	*Yes*
D_4	*Rain*	*Mild*	*High*	*Weak*	*Yes*
D_5	*Rain*	*Cool*	*Normal*	*Weak*	*Yes*
D_6	*Rain*	*Cool*	*Normal*	*Strong*	*No*
D_7	*Overcast*	*Cool*	*Normal*	*Strong*	*Yes*
D_8	*Sunny*	*Mild*	*High*	*Weak*	*No*
D_9	*Sunny*	*Cool*	*Normal*	*Weak*	*Yes*
D_{10}	*Rain*	*Mild*	*Normal*	*Weak*	*Yes*
D_{11}	*Sunny*	*Mild*	*Normal*	*Strong*	*Yes*
D_{12}	*Overcast*	*Mild*	*High*	*Strong*	*Yes*
D_{13}	*Overcast*	*Hot*	*Normal*	*Weak*	*Yes*
D_{14}	*Rain*	*Mild*	*High*	*Strong*	*No*

- Lets calculate all the above required probabilities:

$P(Yes) = \frac{9}{14}$ $\quad$ $P(No) = \frac{5}{14}$

$P(Sunny|Yes) = \frac{2}{9}$ $\quad$ $P(Sunny|No) = \frac{3}{5}$

$P(Cool|Yes) = \frac{3}{9}$ $\quad$ $P(Cool|No) = \frac{1}{5}$

$P(High|Yes) = \frac{3}{9}$ $\quad$ $P(High|No) = \frac{4}{5}$

$P(Strong|Yes) = \frac{3}{9}$ $\quad$ $P(Strong|No) = \frac{3}{5}$

Naïve Bayes Classifier: Example

- Now, the conditional probabilities for the remaining attributes can be calculated as follows:

$$P(yes)P(sunny|yes)P(cool|yes)P(high|yes)P(strong|yes) = \frac{9}{14}\cdot\frac{2}{9}\cdot\frac{3}{9}\cdot\frac{3}{9}\cdot\frac{3}{9} = 0.0053$$

$$P(no)P(sunny|no)P(cool|no)P(high|no)P(strong|no) \ = \frac{5}{14}\cdot\frac{3}{5}\cdot\frac{1}{5}\cdot\frac{4}{5}\cdot\frac{3}{4} = 0.0206$$

- Thus, the Naïve Bayes classifier assigns the target value ***PlayTennis* = *No*** to this new instance.

- Try for the following instance:

  (*Outlook* = ***Shower***, *Temperature* = *cool*, *Humidity* = *high*, *Wind* = *strong*).

Naïve Bayes Classifier: M-Estimate

- ***Problem***: If one of the conditional probability is zero, then the entire expression becomes zero.

  - More extreme case: if the training examples do not cover many of the attribute values, we may not be able to classify some of the test records.

  - For example, if P(Marital Status= Divorced| No) is zero instead of 1/7, then a record with attribute set X=(Home=Yes, Status=Divorced, Income=\$120K) has the following class conditional probabilities:

$$P(X|No) = \frac{3}{7} \times 0 \times 0.0072 = 0$$
$$P(X|Yes) = 0 \times \frac{1}{3} \times 1.2 \times 10^{-9} = 0$$

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

The Naïve Bayes classifier will not be able to classify the record.

- This problem can be addressed by using *m*-estimate approach.

Naïve Bayes Classifier: M-Estimate

- Probability estimation using the *m-estimate* defined as follows:

$$m - \text{estimate} = P(X_i = x_i | Y = y_k) = \frac{N_i + mp}{N_k + m}$$

  - N_i is number of instances having attribute value $X_i = x_i$ and belongs to class y_k.
  - N_k be the total number of training instances for class y_k.
  - p is our prior estimate of the probability we wish to determine.
  - m is a constant called the *equivalent sample size*, which determines how heavily to weight p relative to the observed data.
  - A typical method for choosing p in the absence of other information is to assume uniform priors; that is, if an attribute has r possible values we set $p = \frac{1}{r}$.

Note:

If $n = 0$, that is, if there is no training set available, then $P(x_i|y_k) = p$, so, this is a different value, in absence of sample value.

Naïve Bayes Classifier: M-Estimate

- As in the previous slide, the conditional probability P(Status = Married/Yes) = 0 because none of the training records for the class has the particular attribute value.

- Using the m-estimate approach with m = 3 and $p = \frac{1}{r} = \frac{1}{3}$.

- The conditional probability is no longer zero:

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

$$P(Status = Married|Yes) = \frac{N_i + mp}{N_k + m} = \frac{0 + 3.\frac{1}{3}}{3 + 3} = \frac{1}{6}$$

Naïve Bayes Classifier: M-Estimate

Example 5: *Consider the same instance of Example* 3, X = (Home = No, Married, Income = 120K)

- If we assume p = ½ for the attribute *Home*, p = 1/3 for the attribute *Marital*, then

$$P(Home = No|No) = \frac{N_i + mp}{N_k + m}$$

$$= \frac{4+3.\frac{1}{2}}{7+3} = 0.55$$

$$P(Home = No|Yes) = \frac{3+3.\frac{1}{2}}{3+3} = 0.75$$

$$P(Married|No) = \frac{4+3.\frac{1}{3}}{7+3} = \frac{5}{10} = 0.5$$

$$P(Married|Yes) = \frac{0+3.\frac{1}{3}}{3+3} = \frac{1}{6}$$

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Naïve Bayes Classifier: M-Estimate

Example 5: *Consider the same instance of Example* 3, X = (Home = No, Married, Income = 120K)

- If we assume p = 1/3 for all attributes of class *Yes* and p = 2/3 for all attributes of class *No*, then

  - $P(X|\text{No}) = P(\text{Home=No}|\text{No}) \times P(\text{Married}|\text{No}) \times P(\text{Income=120K}|\text{No})$

    $$= 0.55 \times 0.5 \times 0.0072 = 0.00198$$

  - $P(X|\text{Yes}) = P(\text{Home=No}|\text{Yes}) \times P(\text{Married}|\text{Yes}) \times P(\text{Income=120K}|\text{Yes})$

    $$= 0.75 \times \frac{1}{6} \times 1.2 \times 10^{-9} = 1.5 \times 10^{-10}$$

- The posterior probability for class *No* and *Yes* are:

  $$P(\text{No}|X) = \alpha \times 7/10 \times 0.00198 = 0.001386\alpha$$
  $$P(\text{Yes}|X) = \alpha \times 3/10 \times 1.5 \times 10^{-10} = 4.5 \times 10^{-11}\alpha.$$

- Since, P(No|X) > P(Yes|X), the instance X should be classified as "No".

❖ Although the classification decision has not changed, the m-estimate approach generally provides a more robust way for estimating probabilities when the number of training examples is small.

Naïve Bayes Classifier: M-Estimate

Example 6: For each successive trial, the second row gives the observed outcome; the third, the relative frequency of *heads*; the last, the *m*-estimate of the probability, assuming $p = 0.5$ and $m = 2$.

Toss number	1	2	3	4	5
Outcome	Heads	Heads	Tails	Heads	Tails
Relative frequency	1.00	1.00	0.67	0.75	0.60
m-estimate	0.67	0.75	0.60	0.67	0.57

$$m - \text{estimate} = P(Head) = \frac{N_i + mp}{N_k + m}$$

- As the number of trials increases, though, the values returned by *m*-estimate and relative frequency tend to converge.
- Therefore, *m*-estimate can be more efficient for the dataset with considerably large sample size.
- What is the effect of *m*? Very large *m* value!!

Naïve Bayes Classifier: M-Estimate

- We should not forget that the *m*-estimate is only as good as the parameters it relies on. If we start from an unrealistic prior estimate, the result can be disappointing. Suppose that $p = 0.9$ and $m = 10$. Then,

$$m - \text{estimate} = P(Head) = \frac{N_i + mp}{N_k + m} = \frac{N_i + 10 \times 0.9}{N_k + 10} = \frac{N_i + 9}{N_k + 10}$$

- When we use this formula to recalculate the values of the Table, we will realize that, after five trials, the probability is estimated as:

$$P(Head) = \frac{N_i + 9}{N_k + 10} = \frac{3 + 9}{5 + 10} = 0.8$$

surely a less plausible value than the one obtained in the case of $p = 0.5$ where we got $P(heads) = 0.57$.

Assignments

Assignment NBC1: Consider the following set of training examples and,

a) Estimate the conditional probabilities for $P(X_1|+)$, $P(X_2|+)$, $P(X_3|+)$, $P(X_1|-)$, $P(X_2|-)$, and $P(X_3|-)$. Consider all the categorical values of X_1, X_2, and X_3.

b) Use the estimate of conditional probabilities given in the previous question to predict the class label for a test sample ($X_1 = 0, X_2 = 1, X_3 = 0$) using the Naïve Bayes approach.

c) Estimate the conditional probabilities using the m-estimate approach, with $p = 1/2$ and $m = 3$.

d) Repeat part (b) using the conditional probabilities given in part (c).

e) Compare the two methods for estimating probabilities. Which method is better and why?

Record	X_1	X_2	X_3	***Class***
1	0	0	0	+
2	0	0	1	-
3	0	1	1	-
4	0	1	1	-
5	0	0	1	+
6	1	0	1	+
7	1	0	1	-
8	1	0	1	-
9	1	1	1	+
10	1	0	1	+

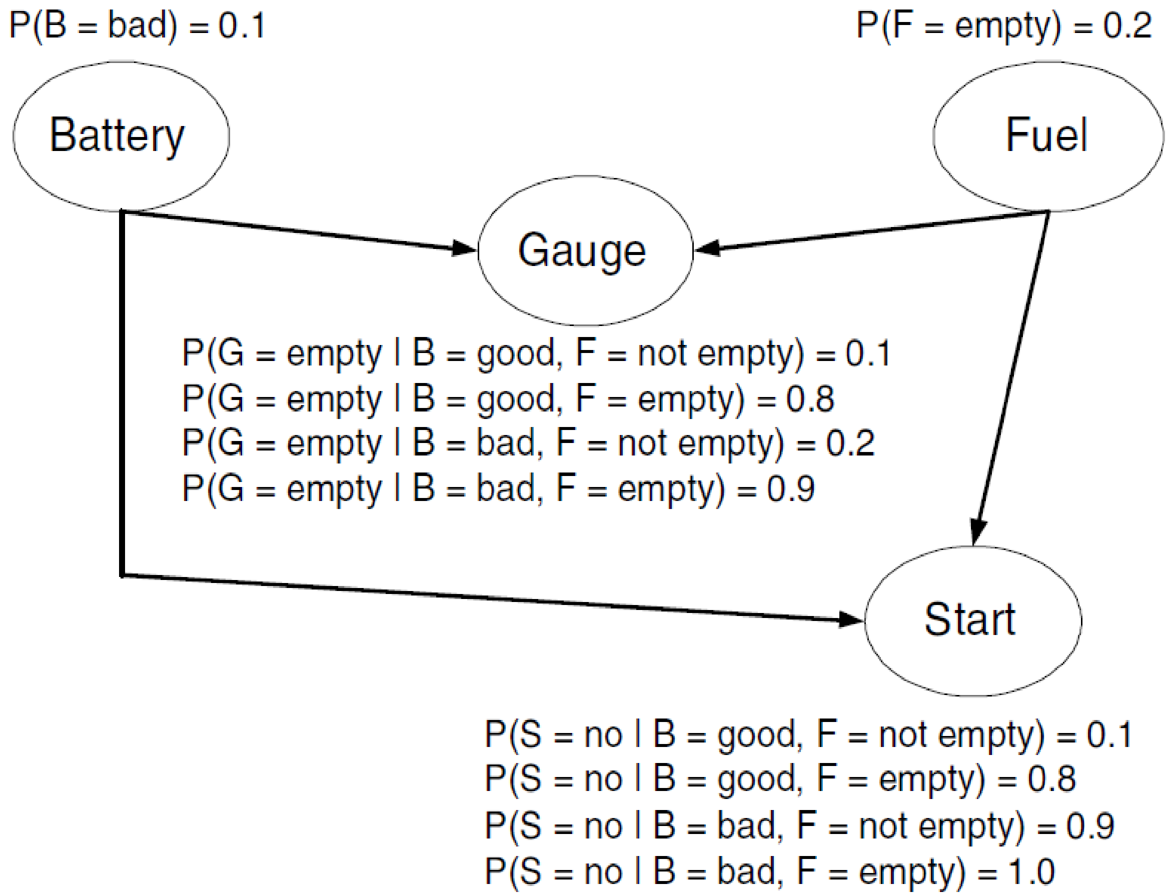
Assignments

Assignment NBC2: Given the Bayesian network shown in the Figure, compute the following probabilities:

a) P(B = good, F = empty, G = empty, S = yes).

b) P(B = bad, F = empty, G = not empty, S = no).

c) Given that the battery is bad, compute the probability that the car will start.



Assignments

Assignment* NBC3:** Draw the Bayesian belief network that represents the conditional independence assumptions of the naive Bayes classifier for the ***PlayTennis problem (refer to slides of Decision Tree). Give the conditional probability table associated with the node ***Wind.***

Books:

1. "Machine Learning" by Tom Mitchell, McGraw Hill.
2. "Introduction to Data Mining" by PN Tang, M Steinbach, V Kumar, Pearson.