Bayes Optimal Classifier,
Gibbs Classifier,
Naïve Bayes Classifier

BAYES OPTIMAL CLASSIFIER

- consider a hypothesis space containing three hypotheses, h1, h2, and h3. Suppose that the posterior probabilities of these hypotheses given the training data are .4, **.3,** and **.3** respectively. Thus, h1 is the MAP hypothesis.
- Suppose a new instance x is encountered, which is classified positive by ***h1,*** but negative by *h2* and h3.
  - Taking all hypotheses into account, the probability that x is positive is .4, and the probability that it is negative is therefore .6.
  - The most probable classification (negative) in this case is different from the classification generated by the MAP hypothesis.

In general, the most probable classification of the new instance is obtained by combining the predictions of all hypotheses, weighted by their posterior probabilities. If the possible classification of the new example can take on any value v_j from some set V, then the probability $P(v_j|D)$ that the correct classification for the new instance is v_j, is just

$$P(v_j|D) = \sum_{h_i \in H} P(v_j|h_i)P(h_i|D)$$

The optimal classification of the new instance is the value v_j, for which $P(v_j|D)$ is maximum.

Bayes optimal classification:

$$\operatorname*{argmax}_{v_j \in V} \sum_{h_i \in H} P(v_j|h_i)P(h_i|D)$$

Example

To illustrate in terms of the above example, the set of possible classifications of the new instance is $V = \{\oplus, \ominus\}$, and

$$P(h_1|D) = .4,\ P(\ominus|h_1) = 0,\ P(\oplus|h_1) = 1$$

$$P(h_2|D) = .3,\ P(\ominus|h_2) = 1,\ P(\oplus|h_2) = 0$$

$$P(h_3|D) = .3,\ P(\ominus|h_3) = 1,\ P(\oplus|h_3) = 0$$

$$\sum_{h_i \in H} P(\oplus|h_i)P(h_i|D) = .4$$

therefore

$$\sum_{h_i \in H} P(\ominus|h_i)P(h_i|D) = .6$$

$$\underset{v_j \in \{\oplus,\ominus\}}{\operatorname{argmax}} \sum_{h_i \in H} P(v_j|h_i)P(h_i|D) = .6$$

GIBBS ALGORITHM

Bayes optimal classifier provides best result, but can be expensive if many hypotheses.

Gibbs algorithm:

1. Choose one hypothesis at random, according to $P(h|D)$
2. Use this to classify new instance

Surprising fact: Assume target concepts are drawn at random from H according to priors on H. Then:

$$E[error_{Gibbs}] \leq 2E[error_{BayesOptimal}]$$

Suppose correct, uniform prior distribution over H, then

- Pick any hypothesis from VS, with uniform probability
- Its expected error no worse than twice Bayes optimal

NAIVE BAYES CLASSIFIER

- One highly practical Bayesian learning method is the naive Bayes learner, often called the ***naive Bayes classifier.***
- In some domains its performance has been shown to be comparable to that of neural network and decision tree learning.
- The naive Bayes classifier applies to learning tasks where each instance ***x*** is described by a conjunction of attribute values and where the target function f ***(x)*** can take on any value from some finite set V.
- A set of training examples of the target function is provided, and a new instance is presented, described by the tuple of attribute values ***(al, a2.. .a,).***
- The learner is asked to predict the target value, or classification, for this new instance.
- Successful Applications: Diagnosis, and classifying Text Document.

The Bayesian approach to classifying the new instance is to assign the most probable target value, v_{MAP}, given the attribute values $\langle a_1, a_2 \ldots a_n \rangle$ that describe the instance.

$$v_{MAP} = \operatorname*{argmax}_{v_j \in V} P(v_j | a_1, a_2 \ldots a_n)$$

We can use Bayes theorem to rewrite this expression as

$$\begin{aligned} v_{MAP} &= \operatorname*{argmax}_{v_j \in V} \frac{P(a_1, a_2 \ldots a_n | v_j) P(v_j)}{P(a_1, a_2 \ldots a_n)} \\ &= \operatorname*{argmax}_{v_j \in V} P(a_1, a_2 \ldots a_n | v_j) P(v_j) \end{aligned}$$

Naive Bayes classifier:

$$v_{NB} = \operatorname*{argmax}_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$$

where v_{NB} denotes the target value output by the naive Bayes classifier.

Naive Bayes assumption:

$$P(a_1, a_2 \ldots a_n | v_j) = \prod_i P(a_i | v_j)$$

Naive Bayes Algorithm

Naive_Bayes_Learn($examples$)

For each target value v_j

$\hat{P}(v_j) \leftarrow$ estimate $P(v_j)$

For each attribute value a_i of each attribute a

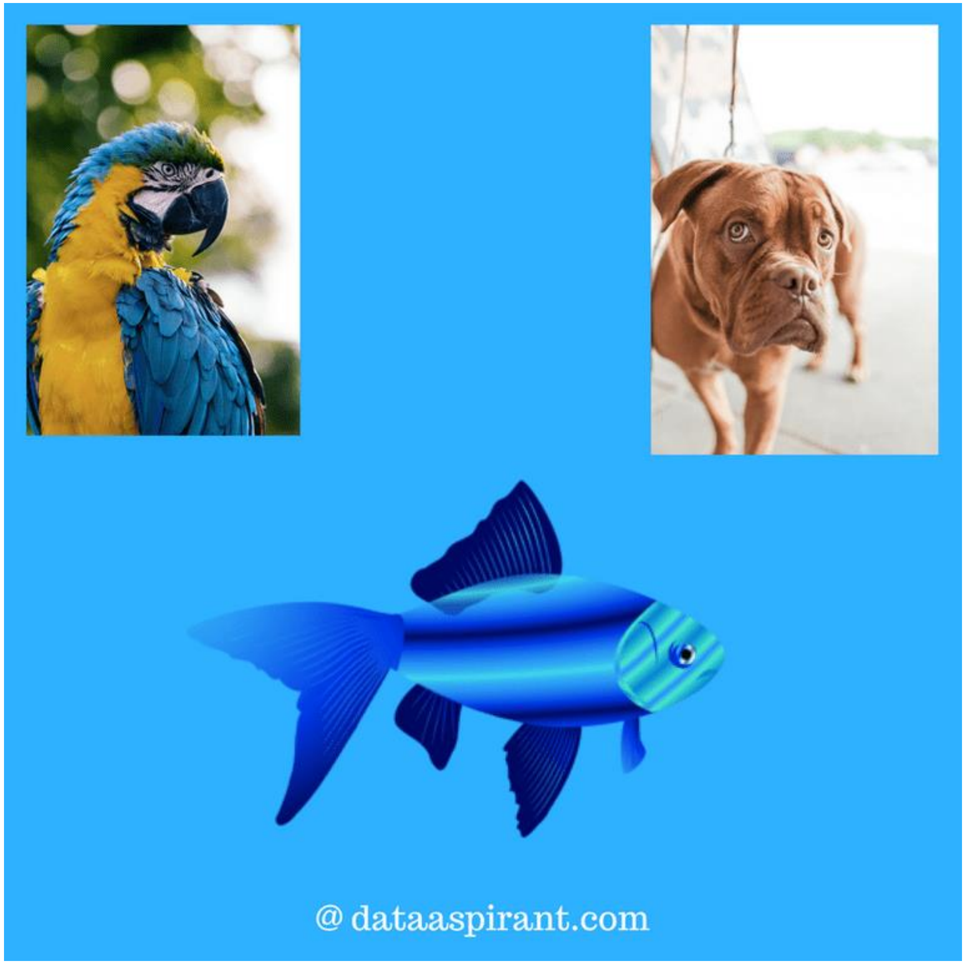
$\hat{P}(a_i|v_j) \leftarrow$ estimate $P(a_i|v_j)$

Classify_New_Instance(x)

$$v_{NB} = \underset{v_j \in V}{\operatorname{argmax}} \hat{P}(v_j) \prod_{a_i \in x} \hat{P}(a_i|v_j)$$

Example of Naïve Bayes Classifier

- **Dataset with 3 classes:**
  - **Parrot**
  - **Dog**
  - **Fish**
- **4 features:**
  - **Swim**
  - **Wings**
  - **Green Color**
  - **Dangerous Teeth**



Example of Naïve Bayes Classifier

Swim	Wings	Green Color	Dangerous Teeth	Animal Type
50	500/500	400/500	0	Parrot
450/500	0	0	500/500	Dog
500/500	0	100/500	50/500	Fish

- **Table shows a frequency table of dataset.**

Example of Naïve Bayes Classifier

- **Considering the dataset predict the type of animal if:**
  - **Swim=True**
  - **Wings=False**
  - **Green=True**
  - **Teeth=False**

- Use naïve bayes approach for multiple evidances:

$$P(H|E_1, E_2 \dots E_N) = \frac{P(E_1|H) * P(E_2|H) \dots * P(E_N|H) * P(H)}{P(E_1, E_2, \dots E_N)}$$

Example of Naïve Bayes Classifier

- **Use naïve bayes approach for multiple evidances:**

$$P(H|E_1, E_2 \dots EN) = \frac{P(E_1|H) * P(E_2|H) \dots * P(E_N|H) * P(H)}{P(E_1, E_2, \dots EN)}$$

Example of Naïve Bayes Classifier

- **For hypothesis testing for animal to be a dog:**

P(Dog|Swim , Green)=P(Swim|Dog)*P(Green|Dog)*P(Dog)/P(Swim, Green)

=0.9*0*0.333/P(Swim, Green)

=0

Example of Naïve Bayes Classifier

- **For hypothesis testing for animal to be a Parrot:**

P(Parrot|Swim , Green)=P(Swim|Parrot)*P(Green|Parrot)*P(Parrot)/ P(Swim, Green)

=0.1*0.8*0.333/P(Swim, Green)

=0.0264/P(Swim, Green)

Example of Naïve Bayes Classifier

- **For hypothesis testing for animal to be a Fish:**

P(Fish|Swim , Green)=P(Swim|Fish)*P(Green|Fish)*P(Fish)/P(Swim, Green)

=1*0.2*0.333/P(Swim, Green)

=0.0666/P(Swim, Green)

Example of Naïve Bayes Classifier

- **For all the three cases, denominator is same.**
- **The value of P(Fish|Swim, Green) is greater than P(Parrot|Swim, Green).**
- **Thus, Class of animal is Fish.**

Exercise

- **Predict the class of animal if:**
  - **Swim=True**
  - **Wings=False**
  - **Green=True**
  - **Teeth=True**

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Consider *PlayTennis* again, and new instance

$\langle Outlk = sun, Temp = cool, Humid = high, Wind = stron$

Want to compute:

$$v_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod_i P(a_i|v_j)$$

$P(y)\ P(sun|y)\ P(cool|y)\ P(high|y)\ P(strong|y) =$

$P(n)\ P(sun|n)\ P(cool|n)\ P(high|n)\ P(strong|n) =$

$$\rightarrow v_{NB} = n$$

Thank You!!!