Naïve Bayes

ECE 4200, 2/5/2020

The classification algorithm

Given a new data $\vec{X}_{n+1}$ predict its label:

$$\arg\max_{y} \underbrace{p(y)p(\vec{X}_{n+1}|y)}.$$

We need to estimate these probabilities.

Naïve Bayes Assumption

$$\bar{X} = \left(\bar{X}^1, \bar{X}^1, \dots, \bar{X}^d\right)$$

Assumption: For any label y and any feature $\bar{X}$

$$p(\bar{X}|y) = \underline{p(\bar{X}^1|y)}\,\underline{p(\bar{X}^2|y)} \dots \underline{p\left(\bar{X}^d\middle|y\right)}$$

Given the label, the features are independent.

Naïve Bayes (NB) Assumption

Example:

Say $\bar{X} = (S, H, N, W)$, and the label $y = No$, then NB:

$$p\ ((Sunny, High, Normal, Weak)|No)$$
$$= p(Sunny|No)p(High|No)p(Normal|No)p(Weak|No)$$

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

TABLE 3.2
Training examples for the target concept *PlayTennis*.

Estimating Probabilities

How to estimate conditional probabilities of features?

attribute.

- CAN consider one ~~feature~~ at a time ---

Consider $\boxed{\text{Outlook}}$. Need to estimate:

$$p(Sunny|No), p(Sunny|Yes),$$
$$p(Overcast|No), p(Overcast|Yes)$$
$$p(Rain|No), p(Rain|Yes)$$

Estimating Probabilities

$$\boxed{p(Sunny|No), p(Overcast|No), p(Rain|No)}$$

Look at all the *No* days: $\boxed{\text{3 Sunny, 2 Rain, 0 overcast.}}$

What will maximum likelihood do?

$p(\text{Sunny}|\text{No}) = \frac{3}{5}$, $p(\text{Rain}|\text{No}) = \frac{2}{5}$, $p(\text{Overcast}|\text{No}) = 0$

$p(\text{Sunny}|\text{Yes}) = \frac{\ }{9}$, $P(\text{Rain}|\text{Yes}) = \frac{\ }{9}$, $p(\text{Ov}|\text{Yes}) = \frac{4}{9}$

Ex:- $Pr(\text{Weak}|\text{No})$ $P(\text{Strong}|\text{No})$

Estimating Probabilities

$p(Sunny|No), p(Overcast|No), p(Rain|No)$

Look at all the *No* days: 3 Sunny, 2 Rain, 0 overcast.

What will maximum likelihood do?

Estimating Probabilities

Zero probabilities can be problem for these ML systems.

Need to take care of it … how?

Laplace: Probability that the sun will rise tomorrow?

$$p(\text{Overcast} \mid \text{No}) = \underline{\underline{0}}.$$

LAPLACE :- What is Prob(sun rise tomorrow) ?

1	1		
5000×365	0	$\frac{1}{1+5000\times 365}$,	$\frac{5000\times 365}{1+5000\times 365}$.
Yes	No		

Estimating Probabilities

Add β estimators.

Outlook, Label, No,

$\boxed{\underset{S}{3+\beta} \quad \underset{R}{2+\beta} \quad \underset{\text{overcast}}{0+\beta}}$ $\boxed{5+3\beta}$

Suppose an attribute has k possibilities. Do a **smoothing**.

$\boxed{\hat{p}_\beta(Sunny|No) = \frac{3+\beta}{5+3\beta}}$, and so on …

Interpretation, Bayesian … Or not?

We then discussed what happens with changing beta.

Prediction

Given a new $\boxed{\bar{X}_{n+1}:}$

$$\arg\max_{y} \boxed{p(y)p(\bar{X}_{n+1}^{1}|y)(\bar{X}_{n+1}^{2}|y)\ldots p(\bar{X}_{n+1}^{d}|y)}$$

Each of these we have modeled already! We now use these for prediction.

(Sunny, High, Normal, Weak)

$y = No$:- $\hat{p}(No) \cdot \hat{p}(S|No) \cdot \hat{p}(H|No) \cdot \hat{p}(N|No) \cdot \hat{p}(W|No)$

$y = Yes$ - $\hat{p}(Yes) \cdot \hat{p}(S|Y)\ \hat{p}(H|Yes)\ \hat{p}(N|Yes) \cdot \hat{p}(Weak|Yes)$

$\hat{p}(No) = \frac{5}{14}$, $\hat{p}(Y) = \frac{9}{14}$.

Gaussian Naïve Bayes

Continuous random variables.

$$\arg\max_y p(y) \cdot p(\bar{x}^1|y) \ldots p(\bar{x}^d|y)$$

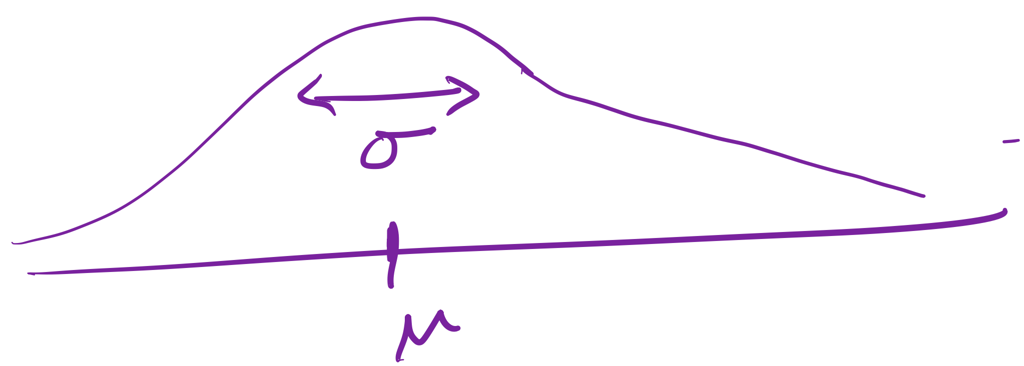
On top of NB assumption, also assume each feature is Gaussian, conditioned on the label!

$$p(\bar{x}|y) = \boxed{p(\bar{x}^1|y)} \cdot p(\bar{x}^2|y) \ldots p(\bar{x}^d|y)$$

How to derive …?

For a label 'y', $p(\bar{x}^1|y)$ is Gaussian.

$$N(\mu, \sigma^2)$$



GNB Assumption

- You have the Naïve Bayes Assumption
  - Features are conditionally independent given labels
- On top of it, assume each feature is distributed according to a Gaussian random variable.
- Gaussian Random variable with mean μ, and variance σ^2:

$$\boxed{p_{\mu,\sigma^2}(x) = \frac{1}{(2\pi\sigma^2)^{0.5}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).}$$

$\boxed{p(\bar{X}^i \mid y)}$ $\rightarrow$ Look at i^{th} attribut for all examples w/ label y!

Estimating conditional mean

For each label y, and each attribute j:

- The goal is to characterize $p(\overrightarrow{X^j}|y)$, which is assumed to be a Gaussian.
- So we need to specify only the mean and variances:

$p(\vec{X}^j|y)$

$$Mean\ Estimate: \mu_{j|y} = \frac{\sum_{i:y_i=y} \overrightarrow{X_i^j}}{\#\ examples\ with\ label\ y}$$

sample mean of the jth attribute for examples with label y

$Var(X) = \mathbb{E}\left[(X - E(X))^2\right]$

Estimating Conditional Variance

For each label y, and each attribute:

- The goal is to characterize $p(\overrightarrow{X^j}|y)$, which is assumed to be a Gaussian.
- So we need to specify only the mean and variances:

$$\boxed{Variance\ Estimate: \sigma_{j|y}^2 = \frac{\sum_{i:y_i=y}\left(\overrightarrow{X_i^j} - \mu_{j|y}\right)^2}{\#\ examples\ with\ label\ y}}$$

Sample variance of the jth for all examples with label y