Naïve Bayes

ECE 4200, 2/5/2020

The classification algorithm

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

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

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

Given the label, the features are independent.

Learning Phase

$$D = \{(\bar{X}_1, y_1), (\bar{X}_2, y_2), \ldots, (\bar{X}_n, y_n)\}$$
$$\bar{X}_i = \left(\bar{X}_i{}^1, \bar{X}_i{}^2, \ldots, \bar{X}_i{}^d\right)$$

- Estimate $p(y)$, which maximizes

$$\prod_{j=1,2,\ldots,n} p(y_j)$$

- Estimate $p\left(\bar{X}^i \middle| y\right)$ for all i, which maximizes

$$\prod_{j=1,2,\ldots,n} p\left(\bar{X}_j{}^i \middle| y\right)$$

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

$$p(Outlook|No)$$

$$p(Outlook = Sunny|No)$$
$$p(Outlook = Overcast|No)$$
$$p(Outlook = Rain|No)$$

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

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

Prediction Phase

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

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

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

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.
- Each $p\left(\bar{X}^i \middle| y\right)$ is a Gaussian Random variable with mean $\mu_{i|y}$, and variance $\sigma_{i|y}^2$:

$$p\left(\bar{X}^i \middle| y\right) = \frac{1}{\left(2\pi\sigma_{i|y}^2\right)^{0.5}} \exp\left(-\frac{\left(\bar{X}^i - \mu_{i|y}\right)^2}{2\sigma_{i|y}^2}\right).$$

Learning Phase

$$D = \{(\bar{X}_1, y_1), (\bar{X}_2, y_2), \dots, (\bar{X}_n, y_n)\}$$
$$\bar{X}_i = \left(\bar{X}_i^{\ 1}, \bar{X}_i^{\ 2}, \dots, \bar{X}_i^{\ d}\right)$$

- Estimate $p(y)$, which maximizes
$$\prod_{j=1,2,\dots,n} p(y_j)$$
- Estimate $p(\bar{X}^i | y)$ for all i, which maximizes
$$\prod_{j=1,2,\dots,n} p\left(\bar{X}_j^{\ i} \middle| y\right)$$

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:

$$\arg\max_{\mu_{i|y},\sigma^2_{i|y}} \prod_{j=1,\ldots,n} \frac{1}{\left(2\pi\sigma^2_{i|y_i}\right)^{0.5}} \exp\left(-\frac{\left(\bar{X}_j{}^i - \mu_{i|y_j}\right)^2}{2\sigma^2_{i|y_j}}\right).$$

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:

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

sample mean of the jth attribute for examples with label y

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:

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

Sample variance of the jth for all examples with label y

Question

If you have 3 possible labels and 4 possible features, how many $\mu_{j|y}$'s will you get? What about $\sigma_{j|y}^2$?

Prediction Phase

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

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

$$\arg\max_{y} p(y) \prod_{i=1,\dots,d} \frac{1}{\left(2\pi\sigma_{i|y}^{2}\right)^{0.5}} \exp\left(-\frac{\left(\bar{X}_{n+1}{}^{i} - \mu_{i|y}\right)^{2}}{2\sigma_{i|y}^{2}}\right).$$