CMPS 460 – Spring 2022

MACHINE LEARNING

Tamer Elsayed

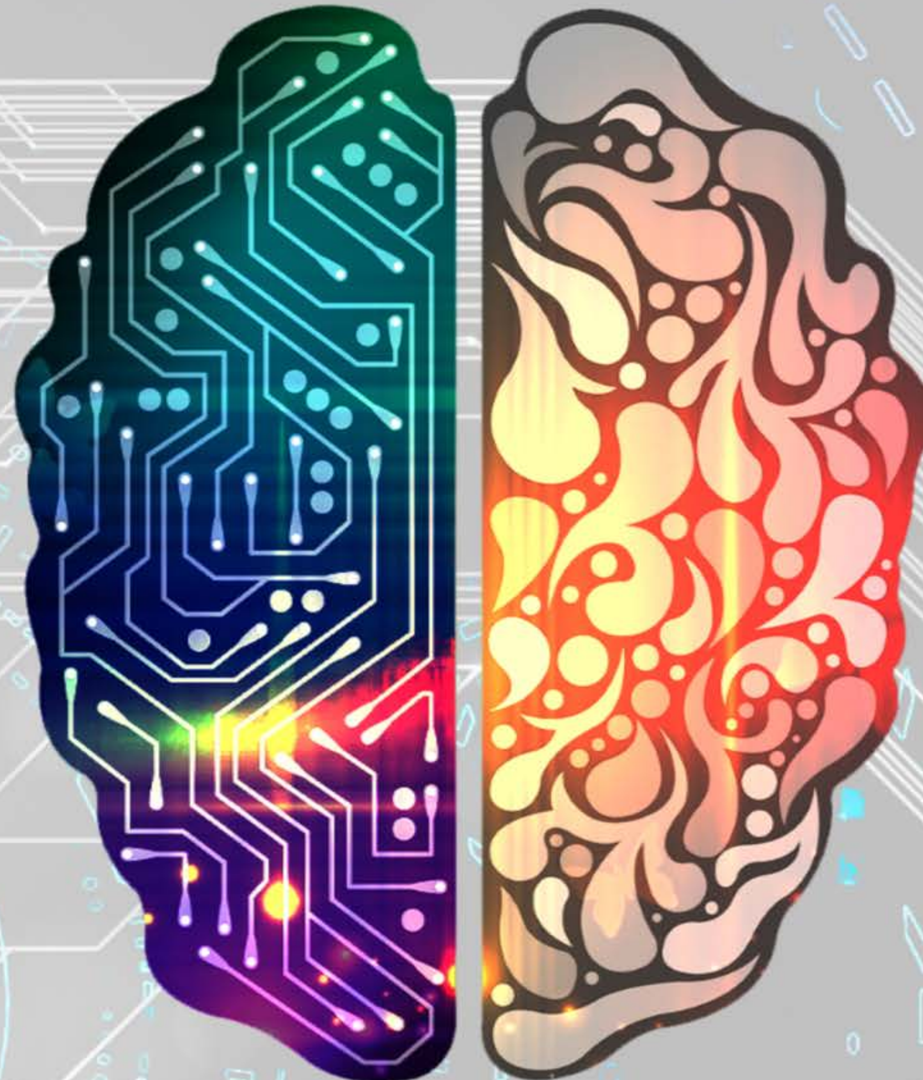


Image hosted by: WittySparks.com | Image source: Pixabay.com

8.b

Probabilistic Modeling: Naïve Bayes

Sec 9.3-9.4

Example Dataset

personal_status	job	housing	savings_status	credit_class
male single	skilled	own	no known savings	good
female div/dep/mar	skilled	own	<100	bad
male single	unskilled resident	own	<100	good
male single	skilled	for free	<100	good
male single	skilled	for free	<100	bad
male single	unskilled resident	for free	no known savings	good
male single	skilled	own	500<=X<1000	good
male single	high qualif/self emp/mgm	rent	<100	good
male div/sep	unskilled resident	own	>=1000	good
male mar/wid	high qualif/self emp/mgm	own	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	own	<100	good
male single	unskilled resident	own	<100	bad
female div/dep/mar	skilled	rent	<100	good
female div/dep/mar	unskilled resident	own	100<=X<500	bad
male single	skilled	own	no known savings	good
male single	skilled	own	no known savings	good
female div/dep/mar	high qualif/self emp/mgm	for free	<100	bad
male single	skilled	own	500<=X<1000	good
male single	skilled	own	<100	good
male single	skilled	rent	500<=X<1000	good
male single	unskilled resident	rent	<100	good
male single	skilled	own	100<=X<500	good
male mar/wid	skilled	own	no known savings	good
male single	unskilled resident	own	<100	good
male mar/wid	unskilled resident	own	<100	good

Naïve Bayes Models

- Probability for a single data point

$$p_{\boldsymbol{\theta}}((y, \boldsymbol{x})) = p_{\boldsymbol{\theta}}(y, x_1, x_2, \ldots, x_D)$$

$$\begin{aligned}
p_{\boldsymbol{\theta}}(x_1, x_2, \ldots, x_D, y) &= p_{\boldsymbol{\theta}}(y) p_{\boldsymbol{\theta}}(x_1 \mid y) p_{\boldsymbol{\theta}}(x_2 \mid y, x_1) p_{\boldsymbol{\theta}}(x_3 \mid y, x_1, x_2) \\
&\quad \cdots p_{\boldsymbol{\theta}}(x_D \mid y, x_1, x_2, \ldots, x_{D-1}) && (9.14) \\
&= p_{\boldsymbol{\theta}}(y) \prod_d p_{\boldsymbol{\theta}}(x_d \mid y, x_1, \ldots, x_{d-1}) && (9.15)
\end{aligned}$$

Challenging!

Naïve Bayes Assumption

Conditional Independence Assumption

Features are independent, conditioned on the label.

$$p(x_d \mid y, x_{d'}) = p(x_d \mid y) \quad , \quad \forall d \neq d'$$

$$p_{\theta}((y,\mathbf{x})) = p_{\theta}(y) \prod_{d} p_{\theta}(x_d \mid y, x_1, \ldots, x_{d-1})$$

$$\Downarrow$$

$$= p_{\theta}(y) \prod_{d} p_{\theta}(x_d \mid y)$$

Model assumption

- Start modeling now!

$$p_{\boldsymbol{\theta}}((y, \boldsymbol{x})) = p_{\boldsymbol{\theta}}(y) \prod_d p_{\boldsymbol{\theta}}(x_d \mid y)$$

e.g., Bernoulli distribution

$$= \left(\theta_0^{[y=+1]} (1 - \theta_0)^{[y=-1]}\right) \prod_d \theta_{(y),d}^{[x_d=1]} (1 - \theta_{(y),d})^{[x_d=0]}$$

$$\hat{\theta}_0 = \frac{1}{N} \sum_n [y_n = +1]$$

$$\hat{\theta}_{(+1),d} = \frac{\sum_n [y_n = +1 \wedge x_{n,d} = 1]}{\sum_n [y_n = +1]}$$

$$\hat{\theta}_{(-1),d} = \frac{\sum_n [y_n = -1 \wedge x_{n,d} = 1]}{\sum_n [y_n = -1]}$$

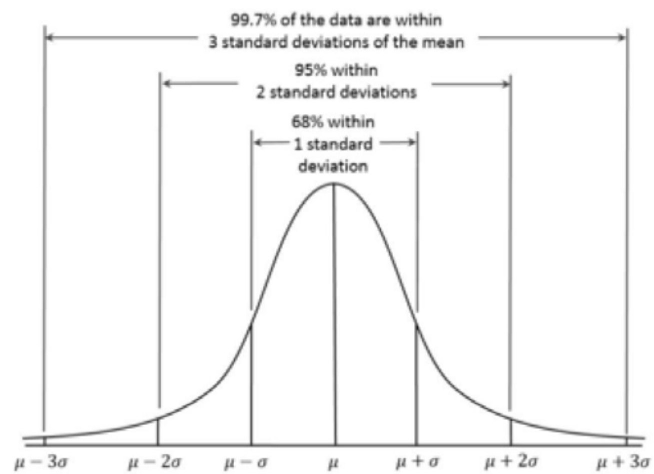
Example Dataset

personal_status	job	housing	savings_status	credit_class
male single	skilled	own	no known savings	good
female div/dep/mar	skilled	own	<100	bad
male single	unskilled resident	own	<100	good
male single	skilled	for free	<100	good
male single	skilled	for free	<100	bad
male single	unskilled resident	for free	no known savings	good
male single	skilled	own	500<=X<1000	good
male single	high qualif/self emp/mgm	rent	<100	good
male div/sep	unskilled resident	own	>=1000	good
male mar/wid	high qualif/self emp/mgm	own	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	own	<100	good
male single	unskilled resident	own	<100	bad
female div/dep/mar	skilled	rent	<100	good
female div/dep/mar	unskilled resident	own	100<=X<500	bad
male single	skilled	own	no known savings	good
male single	skilled	own	no known savings	good
female div/dep/mar	high qualif/self emp/mgm	for free	<100	bad
male single	skilled	own	500<=X<1000	good
male single	skilled	own	<100	good
male single	skilled	rent	500<=X<1000	good
male single	unskilled resident	rent	<100	good
male single	skilled	own	100<=X<500	good
male mar/wid	skilled	own	no known savings	good
male single	unskilled resident	own	<100	good
male mar/wid	unskilled resident	own	<100	good

Choice of distribution …

- In case features are not binary, you need to choose a different model for $p_{\theta}(x_d \mid y)$.

- The choice of distribution is a form of **inductive bias** by which you can inject your knowledge of the problem into the learning algorithm.



Naïve Bayes *Classifier*

Naïve Bayes Classifier

- $\hat{y} = argmax_y \; P(Y = y|X = x)$

$$= argmax_y \frac{P(Y=y)P(X=x|Y=y)}{P(X=x)}$$

$$= argmax_y P(Y = y)P(X = x|Y = y)$$

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

Bayes rule + Conditional independence assumption

Example Training Dataset

- Predict the credit behavior of a credit card applicant from applicant's attributes:
  - Personal status
  - Job type
  - Housing type
  - Savings amount

personal_status	job	housing	savings_status	credit_class
male single	skilled	own	no known savings	good
female div/dep/mar	skilled	own	<100	bad
male single	unskilled resident	own	<100	good
male single	skilled	for free	<100	good
male single	skilled	for free	<100	bad
male single	unskilled resident	for free	no known savings	good
male single	skilled	own	500<=X<1000	good
male single	high qualif/self emp/mgm	rent	<100	good
male div/sep	unskilled resident	own	>=1000	good
male mar/wid	high qualif/self emp/mgm	own	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	own	<100	good
male single	unskilled resident	own	<100	bad
female div/dep/mar	skilled	rent	<100	good
female div/dep/mar	unskilled resident	own	100<=X<500	bad
male single	skilled	own	no known savings	good
male single	skilled	own	no known savings	good
female div/dep/mar	high qualif/self emp/mgm	for free	<100	bad
male single	skilled	own	500<=X<1000	good
male single	skilled	own	<100	good
male single	skilled	rent	500<=X<1000	good
male single	unskilled resident	rent	<100	good
male single	skilled	own	100<=X<500	good
male mar/wid	skilled	own	no known savings	good
male single	unskilled resident	own	<100	good
male mar/wid	unskilled resident	own	<100	good

Training

- Class labels: {good, bad}
  - P(good)=
  - P(bad)=
- Conditional Probabilities
  - P(own|bad)=
  - P(own|good)=
  - P(rent|bad)=
  - P(rent|good)=
  - ... and so on

personal_status	job	housing	savings_status	credit_class
male single	skilled	own	no known savings	good
female div/dep/mar	skilled	own	<100	bad
male single	unskilled resident	own	<100	good
male single	skilled	for free	<100	good
male single	skilled	for free	<100	bad
male single	unskilled resident	for free	no known savings	good
male single	skilled	own	500<=X<1000	good
male single	high qualif/self emp/mgm	rent	<100	good
male div/sep	unskilled resident	own	>=1000	good
male mar/wid	high qualif/self emp/mgm	own	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	own	<100	good
male single	unskilled resident	own	<100	bad
female div/dep/mar	skilled	rent	<100	good
female div/dep/mar	unskilled resident	own	100<=X<500	bad
male single	skilled	own	no known savings	good
male single	skilled	own	no known savings	good
female div/dep/mar	high qualif/self emp/mgm	for free	<100	bad
male single	skilled	own	500<=X<1000	good
male single	skilled	own	<100	good
male single	skilled	rent	500<=X<1000	good
male single	unskilled resident	rent	<100	good
male single	skilled	own	100<=X<500	good
male mar/wid	skilled	own	no known savings	good
male single	unskilled resident	own	<100	good
male mar/wid	unskilled resident	own	<100	good

Training

- Class labels: {good, bad}
  - P(good)=19/27=0.7
  - P(bad)=8/27=0.3
- Conditional Probabilities
  - P(own|bad)=4/8=0.5
  - P(own|good)=13/19=0.68
  - P(rent|bad)=2/8=0.25
  - P(rent|good)=4/19=0.21
  - … and so on

personal_status	job	housing	savings_status	credit_class
male single	skilled	own	no known savings	good
female div/dep/mar	skilled	own	<100	bad
male single	unskilled resident	own	<100	good
male single	skilled	for free	<100	good
male single	skilled	for free	<100	bad
male single	unskilled resident	for free	no known savings	good
male single	skilled	own	500<=X<1000	good
male single	high qualif/self emp/mgm	rent	<100	good
male div/sep	unskilled resident	own	>=1000	good
male mar/wid	high qualif/self emp/mgm	own	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	own	<100	good
male single	unskilled resident	own	<100	bad
female div/dep/mar	skilled	rent	<100	good
female div/dep/mar	unskilled resident	own	100<=X<500	bad
male single	skilled	own	no known savings	good
male single	skilled	own	no known savings	good
female div/dep/mar	high qualif/self emp/mgm	for free	<100	bad
male single	skilled	own	500<=X<1000	good
male single	skilled	own	<100	good
male single	skilled	rent	500<=X<1000	good
male single	unskilled resident	rent	<100	good
male single	skilled	own	100<=X<500	good
male mar/wid	skilled	own	no known savings	good
male single	unskilled resident	own	<100	good
male mar/wid	unskilled resident	own	<100	good

Training Model

X_j	Y_i	$P(X_j \mid Y_i)$
female single	good	0.28
female single	bad	0.36
……..	……	……
own	good	0.75
own	bad	0.62
……..	……	……
self emp	good	0.14
self emp	bad	0.17
……..	……	……
savings>1K	good	0.06
savings>1K	bad	0.02
……..	……	……

Prediction

- Given applicant attributes of

  A= {female single,
  owns home,
  self-employed,
  savings > $1000}

Assume this is part of the trained model

X_j	Y_i	$P(X_j \mid Y_i)$
female single	good	0.28
female single	bad	0.36
own	good	0.75
own	bad	0.62
self emp	good	0.14
self emp	bad	0.17
savings>1K	good	0.06
savings>1K	bad	0.02

$P(good|A) \sim (0.28*0.75*0.14*0.06)*0.7 = 0.0012$

$P(bad|A) \sim (0.36*0.62*0.17*0.02)*0.3 = 0.0002$

- Since P(good|A) > (bad|A), assign the applicant the label "good" credit

Hands-on Exercise

Consider the following training set.

- Learn a NB Model

- Predict the label for X = (c, T, 0)

Training Set

X1	X2	X3	Y
b	T	-1	0
b	T	2	0
c	F	0	1
a	T	0	1
b	F	-1	1
a	T	2	1
c	T	-1	0
b	F	-1	1
c	F	2	1
a	T	0	0

Training a Naïve Bayes classifier

Assume discrete X_i and Y

TrainNaïveBayes (Data)

for each value y_k of Y

estimate $\pi_k = P(Y = y_k)$ $\qquad \dfrac{N(Y = y_k)}{N}$

for each feature X_i

for each value x_{ij} of X_i

estimate $\theta_{ijk} = P(X_i = x_{ij} | Y = y_k)$

$$\frac{N(X_i = x_{ij} \text{ and } Y = y_k)}{N(Y = y_k)}$$

Unobserved Feature/Class Pairs!

What happens if
a feature value didn't co-occur with a class?

- Zero probabilities due to unobserved (or rare) events!

Solution: Smoothing!

- **Smoothing:** adjusting each probability by a small value
- **Add-one Smoothing**

$$P(a) = \frac{N(a) + 1}{\sum_{a'}(N(a') + 1)}$$