COMP47460 Tutorial

Naïve Bayes Classifiers

Aonghus Lawlor
Derek Greene

School of Computer Science
Autumn 2016

Tutorial Q1(a)

a) Construct the contingency table of conditional and prior probabilities that would be used by Naïve Bayes to build a classifier for this dataset.

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
1	Moderate	Moderate	Warm	Light	Some	Yes
2	Light	Moderate	Warm	Moderate	None	No
3	Moderate	Moderate	Cold	Gale	None	No
4	Moderate	Moderate	Warm	Light	None	Yes
5	Moderate	Light	Cold	Light	Some	No
6	Heavy	Light	Cold	Moderate	Some	Yes
7	Light	Light	Cold	Moderate	Some	No
8	Moderate	Moderate	Cold	Gale	Some	No
9	Heavy	Heavy	Warm	Moderate	None	Yes
10	Light	Light	Cold	Light	Some	No

Tutorial Q1(a)

Construct full contingency table for all features on both classes:

Swimming	Yes	No
Rain Recently=light	0/4	3/6
Rain Recently=moderate	2/4	3/6
Rain Recently=heavy	2/4	0/6
Rain Today=light	1/4	3/6
Rain Today=moderate	2/4	3/6
Rain Today=heavy	1/4	0/6
Temp=Cold	1/4	5/6
Temp=Warm	3/4	1/6
Wind=Light	2/4	2/6
Wind=Moderate	2/4	2/6
Wind=Gale	0/4	2/6
Sunshine=Some	2/4	4/6
Sunshine=None	2/4	2/6
Class Probabilities (Priors)	4/10	6/10

Tutorial Q1(b)

b) Based on the contingency table, classify the two new examples below using Naïve Bayes.

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
X1	Heavy	Moderate	Warm	Light	Some	???

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
X2	Light	Moderate	Warm	Light	Some	???

Naïve Bayes classification steps:

1. Calculate probability of input having class *Yes*
2. Calculate probability of input having class *No*
3. Normalise probabilities (optional)

Tutorial Q1(b): Input X1

Test input example for hypothesis 1: *Swimming=Yes*

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
1	Moderate	Moderate	Warm	Light	Some	Yes
2	Light	Moderate	Warm	Moderate	None	No
3	Moderate	Moderate	Cold	Gale	None	No
4	Moderate	Moderate	Warm	Light	None	Yes
5	Moderate	Light	Cold	Light	Some	No
6	Heavy	Light	Cold	Moderate	Some	Yes
7	Light	Light	Cold	Moderate	Some	No
8	Moderate	Moderate	Cold	Gale	Some	No
9	Heavy	Heavy	Warm	Moderate	None	Yes
10	Light	Light	Cold	Light	Some	No

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
X1	Heavy	Moderate	Warm	Light	Some	???

Apply NB for *Swimming=Yes* by calculating product of probabilities for input's feature values and class probability:

$$P = (2/4 \times 2/4 \times 3/4 \times 2/4 \times 2/4) \times 4/10$$

$$P = 0.01875$$

Class Probability

P(Yes) = 4/10

Feature: Rain Recently

P(L_RR|Yes) = 0/4

P(M_RR|Yes) = 2/4

P(H_RR|Yes) = 2/4

Feature: Rain Today

P(L_RT|Yes) = 1/4

P(M_RT|Yes) = 2/4

P(H_RT|Yes) = 1/4

Feature: Temp

P(C_T|Yes) = 1/4

P(W_T|Yes) = 3/4

Feature: Wind

P(L_W|Yes) = 2/4

P(M_W|Yes) = 2/4

P(G_W|Yes) = 0/4

Feature: Sunshine

P(S_S|Yes) = 2/4

P(N_S|Yes) = 2/4

Tutorial Q1(b): Input X1

Test input example for hypothesis 2: *Swimming=No*

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
1	Moderate	Moderate	Warm	Light	Some	Yes
2	Light	Moderate	Warm	Moderate	None	No
3	Moderate	Moderate	Cold	Gale	None	No
4	Moderate	Moderate	Warm	Light	None	Yes
5	Moderate	Light	Cold	Light	Some	No
6	Heavy	Light	Cold	Moderate	Some	Yes
7	Light	Light	Cold	Moderate	Some	No
8	Moderate	Moderate	Cold	Gale	Some	No
9	Heavy	Heavy	Warm	Moderate	None	Yes
10	Light	Light	Cold	Light	Some	No

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
X1	Heavy	Moderate	Warm	Light	Some	???

Apply NB for *Swimming=No* by calculating product of probabilities for input's feature values and class probability:

$$P = (0/6 \times 3/6 \times 1/6 \times 2/6 \times 4/6) \times 6/10$$

$$P = 0$$

Class Probability

$P(No) = 6/10$

Feature: Rain Recently

$P(L\_RR|No) = 3/6$

$P(M\_RR|No) = 3/6$

$P(H\_RR|No) = 0/6$

Feature: Rain Today

$P(L\_RT|No) = 3/6$

$P(M\_RT|No) = 3/6$

$P(H\_RT|No) = 0/6$

Feature: Temp

$P(C\_T|No) = 5/6$

$P(W\_T|No) = 1/6$

Feature: Wind

$P(L\_W|No) = 2/6$

$P(M\_W|No) = 2/6$

$P(G\_W|No) = 2/6$

Feature: Sunshine

$P(S\_S|No) = 4/6$

$P(N\_S|No) = 2/6$

Tutorial Q1(b): Input X1

- Calculated probabilities for two hypotheses (class labels):

  Yes `P(Y) = 2/4 x 2/4 x 3/4 x 2/4 x 2/4 x 4/10 = 0.01875`

  No `P(N) = 0/6 x 3/6 x 1/6 x 2/6 x 4/6 x 6/10 = 0`

- Normalise probabilities to sum to 1:

  Yes `P(Y)' = 0.01875/(0.01875+0) = 1.0`

  No `P(N)' = 0`

- Output Prediction: **Swimming = Yes**

Tutorial Q1(b): Input X2

Test input example for hypothesis 1: *Swimming=Yes*

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
1	Moderate	Moderate	Warm	Light	Some	Yes
2	Light	Moderate	Warm	Moderate	None	No
3	Moderate	Moderate	Cold	Gale	None	No
4	Moderate	Moderate	Warm	Light	None	Yes
5	Moderate	Light	Cold	Light	Some	No
6	Heavy	Light	Cold	Moderate	Some	Yes
7	Light	Light	Cold	Moderate	Some	No
8	Moderate	Moderate	Cold	Gale	Some	No
9	Heavy	Heavy	Warm	Moderate	None	Yes
10	Light	Light	Cold	Light	Some	No

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
X2	Light	Moderate	Warm	Light	Some	???

Apply NB for *Swimming=Yes* by calculating product of probabilities for input's feature values and class probability:

$$P = (0/4 \times 2/4 \times 3/4 \times 2/4 \times 2/4) \times 4/10$$

$$P = 0$$

Class Probability

P(Yes) = 4/10

Feature: Rain Recently

P(L_RR|Yes) = 0/4

P(M_RR|Yes) = 2/4

P(H_RR|Yes) = 2/4

Feature: Rain Today

P(L_RT|Yes) = 1/4

P(M_RT|Yes) = 2/4

P(H_RT|Yes) = 1/4

Feature: Temp

P(C_T|Yes) = 1/4

P(W_T|Yes) = 3/4

Feature: Wind

P(L_W|Yes) = 2/4

P(M_W|Yes) = 2/4

P(G_W|Yes) = 0/4

Feature: Sunshine

P(S_S|Yes) = 2/4

P(N_S|Yes) = 2/4

Tutorial Q1(b): Input X2

Test input example for hypothesis 2: *Swimming=No*

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
1	Moderate	Moderate	Warm	Light	Some	Yes
2	Light	Moderate	Warm	Moderate	None	No
3	Moderate	Moderate	Cold	Gale	None	No
4	Moderate	Moderate	Warm	Light	None	Yes
5	Moderate	Light	Cold	Light	Some	No
6	Heavy	Light	Cold	Moderate	Some	Yes
7	Light	Light	Cold	Moderate	Some	No
8	Moderate	Moderate	Cold	Gale	Some	No
9	Heavy	Heavy	Warm	Moderate	None	Yes
10	Light	Light	Cold	Light	Some	No

Example	Rain Recently (RR)	Rain Today (RT)	Temp (T)	Wind (W)	Sunshine (S)	Swimming
X2	Light	Moderate	Warm	Light	Some	???

Apply NB for *Swimming=No* by calculating product of probabilities for input's feature values and class probability:

$$P = (3/6 \times 3/6 \times 1/6 \times 2/6 \times 4/6) \times 6/10$$

$$P = 0.0056$$

Class Probability

P(No) = 6/10

Feature: Rain Recently

P(L_RR|No) = 3/6

P(M_RR|No) = 3/6

P(H_RR|No) = 0/6

Feature: Rain Today

P(L_RT|No) = 3/6

P(M_RT|No) = 3/6

P(H_RT|No) = 0/6

Feature: Temp

P(C_T|No) = 5/6

P(W_T|No) = 1/6

Feature: Wind

P(L_W|No) = 2/6

P(M_W|No) = 2/6

P(G_W|No) = 2/6

Feature: Sunshine

P(S_S|No) = 4/6

P(N_S|No) = 2/6

Tutorial Q1(b): Input X2

- Calculated probabilities for two hypotheses (class labels):

  Yes `P(Y) = (0/4 x 2/4 x 3/4 x 2/4 x 2/4) x 4/10 = 0`

  No `P(N) = (3/6 x 3/6 x 1/6 x 2/6 x 4/6) x 6/10 = 0.0056`

- Normalise probabilities to sum to 1:

  Yes `P(Y)' = 0`

  No `P(N)' = 0.0056/(0.0056+0) = 1.0`

- Output Prediction: **Swimming = No**

Tutorial Q2(a)

a) Provide the contingency table of conditional and prior probabilities that would be used by Naïve Bayes to build a classifier for this dataset.

	Name	Hair	Height	Build	Lotion	Result
1	Sarah	blonde	average	light	no	sunburned
2	Dana	blonde	tall	average	yes	none
3	Alex	brown	short	average	yes	none
4	Annie	blonde	short	average	no	sunburned
5	Emily	red	average	heavy	no	sunburned
6	Pete	brown	tall	heavy	no	none
7	John	brown	average	heavy	no	none
8	Katie	brown	short	light	yes	none

Tutorial Q2(a)

Construct full contingency table for all features on both classes:

Result	Sunburned	None
Hair=blonde	2/3	1/5
Hair=brown	0/3	4/5
Hair=red	1/3	0/5
Height=average	2/3	1/5
Height=tall	0/3	2/5
Height=short	1/3	2/5
Build=light	1/3	1/5
Build=average	1/3	2/5
Build=heavy	1/3	2/5
Lotion=no	3/3	2/5
Lotion=yes	0/3	3/5
Class Probabilities (Priors)	3/8	5/8

Tutorial Q2(b)

Use the contingency table to calculate the Naïve Bayes scores:

	Hair	Height	Build	Lotion	Result
X	blonde	average	heavy	no	???

$$v_{NB} = \arg\max_{v_j \in V} P(v_j) \prod_i P(f_i|v_j)$$

Result	Sunburned	None
Hair=blonde	2/3	1/5
Hair=brown	0/3	4/5
Hair=red	1/3	0/5
Height=average	2/3	1/5
Height=tall	0/3	2/5
Height=short	1/3	2/5
Build=light	1/3	1/5
Build=average	1/3	2/5
Build=heavy	1/3	2/5
Lotion=no	3/3	2/5
Lotion=yes	0/3	3/5
Class Probabilities	3/8	5/8

Calculate raw probabilities for two classes:

```
P(S) = (2/3)*(2/3)*(1/3)*(3/3) * (3/8)
P(S) = 0.056
P(N) = (1/5)*(1/5)*(2/5)*(2/5) * (5/8)
P(N) = 0.004
```

Normalise probabilities:

```
P(S)' = 0.056/(0.056+0.004) = 0.933
P(N)' = 0.004/(0.056+0.004) = 0.067
```

➡ **Output: Sunburned**

Tutorial Q3(a)

a) Calculate the contingency table that would be used by Naïve Bayes to build a classifier using this training data.

Example	Credit History	Debt	Income	Risk
1	bad	low	0to30	high
2	bad	high	30to60	high
3	bad	low	0to30	high
4	unknown	high	30to60	high
5	unknown	high	0to30	high
6	good	high	0to30	high
7	bad	low	over60	medium
8	unknown	low	30to60	medium
9	good	high	30to60	medium
10	unknown	low	over60	low
11	unknown	low	over60	low
12	good	low	over60	low
13	good	high	over60	low
14	good	high	over60	low

Tutorial Q3(a)

a) Calculate the contingency table that would be used by Naïve Bayes to build a classifier using this training data.

Contingency table (probability table) for each of the descriptive features:

Risk	high	medium	low
CH=bad	3/6	1/3	0
CH=unknown	2/6	1/3	2/5
CH=good	1/6	1/3	3/5
Debt=low	2/6	2/3	3/5
Debt=high	4/6	1/3	2/5
Income=0to30	4/6	0	0
Income=30to60	2/6	2/3	0
Income=over60	0	1/3	5/5
Class Probabilities (Priors)	*6/14*	*3/14*	*5/14*

Tutorial Q3(a)

b) Based on the contingency table, predict a risk level for the new loan application X below.

Example	Credit History	Debt	Income	Risk
X	bad	low	30to60	???

Calculate raw probabilities for 3 classes, using contingency table:

```
P(H) = (3/6)*(2/6)*(2/6) * (6/14) = 0.0238
P(M) = (1/3)*(2/3)*(2/3) * (3/14) = 0.0317
P(L) = (0)*(3/5)*(0) * (5/14) = 0
```

Normalise probabilities:

```
P(H)' = 0.0238/(0.0238+0.0317+0) = 0.4288
P(M)' = 0.0317/(0.0238+0.0317+0) = 0.5712
P(L)' = 0
```

➡ **Output: Medium Risk**