Workshop Week 10 - COMP20008 2021

1. Consider the following data set for a binary class problem:

Feature A	Feature B	Class Label
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
F	F	-

We wish to select the feature that best predicts the class label using the χ^2 method.

- Write down the observed and expected contingency tables for feature A
- Calculate the $\chi^2(A, Class)$ value.
- Using the table below, conclude whether feature A is independent of the class label for $p = 0.05$.

df	P = 0.05	P = 0.01	P = 0.001
1	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.35	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
6	12.59	16.81	22.46

- Repeat the process for feature B and decide which feature could be best used for predicting the class label.

Observed table:

A	A=T	A=F	Total
Class=+	4	0	4
Class=-	2	4	6
Total	6	4	10

Expected table:

A	A=T	A=F	Total
Class=+	2.4	1.6	4
Class=-	3.6	2.4	6
Total	6	4	10

$\chi^2(A, Class) = \frac{(4-2.4)^2}{2.4} + \frac{(0-1.6)^2}{1.6} + \frac{(2-3.6)^2}{3.6} + \frac{(4-2.4)^2}{2.4} = 4.44$

Degrees of freedom $= (2-1) \times (2-1) = 1$
Lookup value in table (3.84). Since our calculated χ^2 value is greater than the critical value in the table, conclude A is not independent of Class for $p = 0.05$

For feature B:
Observed table:

B	B=T	B=F	Total
Class=+	3	1	4
Class=-	1	5	6
Total	4	6	10

Expected table:

B	B=T	B=F	Total
Class=+	1.6	2.4	4
Class=-	2.4	3.6	6
Total	4	6	10

$\chi^2(B, Class) = \frac{(3-1.6)^2}{1.6} + \frac{(1-2.4)^2}{2.4} + \frac{(1-2.4)^2}{3.6} + \frac{(5-3.6)^2}{3.6} = 3.40$

Degrees of freedom $= (2-1) \times (2-1) = 1$
Lookup value in table (3.84). Since our calculated χ^2 value is less than the critical value in the table, conclude B is independent of Class for $p = 0.05$

Feature A best predicts class label.