UNIVERSITY OF LEICESTER

Theoretical Homework 2

For DATA MINING AND NEURAL NETWORKS MA4022

Gleb Vorobchuk

Leicester

May 2018

gv@student.le.ac.uk

Task №1.

A data table with the values of two Boolean variables X1,X2, is presented below. Please find the tables of frequencies for each attribute value and for all combinations of their values. Please evaluate the joint probability distribution by the naïve Bayes method. Please evaluate the probability P(X1≥X2) using the table and the Naïve Bayes evaluation. Compare the results.

x1	x2
0	1
0	1
1	1
1	0
1	0
0	0
1	0
0	1
1	0
0	1

x1	frequency	p(x1)
0	5	0.5
1	5	0.5

x2	frequency	p(x2)
0	4	0.4
1	6	0.6

x1	x2	frequency
0	0	0
0	1	5
1	0	4
1	1	1

$p(x_1 = 0) = 0.5$

$p(x_1 = 1) = 0.5$

$p(x_2 = 0) = 0.4$

$p(x_2 = 1) = 0.6$

Assuming x_1 independent x_2:

$$p(x_1 = 0 \cap x_2 = 0) = p(x_1 = 0)p(x_2 = 0) = 0$$

$$p(x_1 = 0 \cap x_2 = 1) = p(x_1 = 0)p(x_2 = 1) = 0.5 \times 0.6 = 0.3$$

$$p(x_1 = 1 \cap x_2 = 0) = p(x_1 = 1)p(x_2 = 0) = 0.5 \times 0.4 = 0.2$$

$$p(x_1 = 1 \cap x_2 = 1) = p(x_1 = 1)p(x_2 = 1) = 0.5 \times 0.6 = 0.3$$

Using table

x1	x2	frequency
0	0	0
0	1	5
1	0	4
1	1	1

$$p(x_1 \geq x_2) = 0 + 0.4 + 0.1 = 0.5$$

Using joint probability distribution:

$p(x_1 \geq x_2) = p(x_1 = 0 \cap x_2 = 0) + p(x_1 = 1 \cap x_2 = 0) + p(x_1 = 1 \cap x_2 = 1) = 0 + 0.2 + 0.3 = 0.5$

Task №2.

Under which condition a general AR(n) process is stationary? For which a the following processes are stationary: X(t)=aX(t-1)+2aX(t-2)+Z(t); X(t)=2aX(t-1)+aX(t- 2)+Z(t)?

Stationary condition for AR(n) process is that the roots of the equation:

$$\phi(L) = 1 - \alpha_1 L - \ldots - \alpha_p L^p = 0$$

Must lie outside the unit circle. Also called causality condition.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$$

$$X_t = L\phi_1 X_{t-1} + L\phi_2 X_{t-2} + Z_t$$

$$(1 - L\phi_1 - L^2\phi_2)X_t = Z_t$$

Stationary conditions:

$$1)\phi_1^2 + \phi_2^2 < 1$$

$$2)\phi_1 + \phi_2 \neq 1$$

$$3)|\phi_1 + \phi_2| < 1$$

So for the first process:

$$X_t = aX_{t-1} + 2aX_{t-2} + Z_t$$

$$\phi_1 = a; \phi_2 = 2a$$

$$(1 - aL - 2aL^2)X_t = Z_t$$

Stationary if $|L| > 1$

$$\frac{-b \pm \sqrt{b^2 + 4ac}}{2a} \implies \frac{a \pm \sqrt{a^2 + 8a}}{4a}$$

Solution:

$$-\frac{1}{2} < a < 0$$

$$0 < a < \frac{1}{3}$$

So for the second process:

$$X_t = 2aX_{t-1} + aX_{t-2} + Z_t$$

$$\phi_1 = 2a; \phi_2 = a$$

$$(1 - 2aL - aL^2)X_t = Z_t$$

Stationary if $|L| > 1$

$$\frac{-b \pm \sqrt{b^2 + 4ac}}{2a} \implies \frac{a \pm \sqrt{a^2 + a}}{a}$$

Solution:

$$-\frac{1}{3} < a < 0$$

$$0 < a < 1$$

Task №3.

Under which condition a general MA(n) process is invertible? For which a the following processes are invertible: X(t)= Z(t) +aZ(t-1)+2aZ(t-2); X(t)=Z(t)+2aZ(t- 1)+aZ(t-2)?

Invertibility condition for MA(n) is that all the roots of the equation:

$$\theta(L) = 0$$

lie outside the unit circle.

$$X_t = Z_t + aZ_{t-1} + 2aZ_{t-2}$$

$$X_t = L\theta_1 X_{t-1} + L\theta_2 X_{t-2} + Z_t$$

$$(1 - L\theta_1 - L^2\theta_2)X_t = Z_t$$

So for the first process:

$$X_t = aX_{t-1} + 2aX_{t-2} + Z_t$$

$$\theta_1 = a; \theta_2 = 2a$$

$$(1 + aL + 2aL^2)X_t = Z_t$$

Invertible if $|L| > 1$

$$\frac{-b \pm \sqrt{b^2 + 4ac}}{2a} \implies \frac{a \pm \sqrt{a^2 - 8a}}{4a}$$

Solution:

$$-\frac{1}{3} < a < 0$$

$$0 < a < \frac{1}{2}$$

So for the second process:

$$X_t = 2aX_{t-1} + aX_{t-2} + Z_t$$

$$\theta_1 = 2a; \theta_2 = a$$

$$(1 + 2aL + aL^2)X_t = Z_t$$

Invertible if $|L| > 1$

$$\frac{-b \pm \sqrt{b^2 + 4ac}}{2a} \implies \frac{a \pm \sqrt{a^2 - a}}{a}$$

Solution:

$$-\frac{1}{3} < a < 0$$

$$0 < a < 1$$

Task №4.

Under which conditions a general ARMA(m,n) process is stationary and invertible? For which a the following process is stationary and invertible X(t)=aX(t-1)+2aX(t-2)+Z(t) +aZ(t-1)+2aZ(t-2)?

Invertibility condition for MA(n) is that all the roots of the equation:

$$\theta(L) = 0$$

lie outside the unit circle.

Stationary condition for AR(n) process is that the roots of the equation:

$$\phi(L) = 1 - \alpha_1 L - \ldots - \alpha_p L^p = 0$$

lie outside the unit circle.

Separate process on two parts AR and MA:

$X_t = aX_{t-1} + 2aX_{t-2} + Z_t$ and $X_t = aX_{t-1} + 2aX_{t-2} + Z_t$

So for the process:

$$X_t = aX_{t-1} + 2aX_{t-2} + Z_t$$

$$\phi_1 = a; \phi_2 = 2a$$

$$(1 - aL - 2aL^2)X_t = Z_t$$

Stationary if $|L| > 1$

$$\frac{-b \pm \sqrt{b^2 + 4ac}}{2a} \implies \frac{a \pm \sqrt{a^2 + 8a}}{4a}$$

Solution:

$$-\frac{1}{2} < a < 0$$

$$0 < a < \frac{1}{3}$$

$$X_t = aX_{t-1} + 2aX_{t-2} + Z_t$$

$$\theta_1 = a; \theta_2 = 2a$$

$$(1 + aL + 2aL^2)X_t = Z_t$$

Invertible if $|L| > 1$

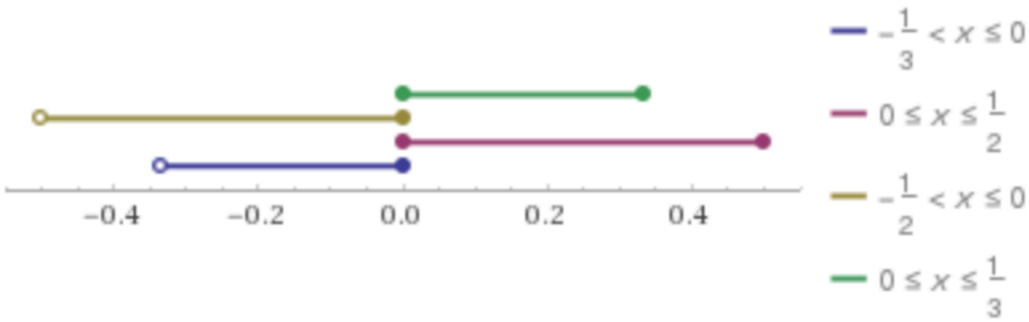
$$\frac{-b \pm \sqrt{b^2 + 4ac}}{2a} \implies \frac{a \pm \sqrt{a^2 - 8a}}{4a}$$

Solution:

$$-\frac{1}{3} < a < 0$$

$$0 < a < \frac{1}{2}$$

Combining solutions:



Finan answer is:

$$-\frac{1}{3} < a < 0$$

$$0 < a < \frac{1}{3}$$

Task №5.

Define autocorrelation function and calculate it for a AR(1) process.

Autocorrelation function or ACF[1]:

$$\rho_p = \frac{cov\{y_t, y_t - k\}}{var\{y_t\}}$$

AR(n)

$$X_t = \alpha_1 X_{t-1} \ldots \alpha_n X_{t-n} + Z_t$$

$$X_t = Z_t + \sum_{i=1}^{P} \alpha_p X_{t-p}$$

AR(1)

$$X_t = \alpha_1 X_{t-1} + Z_t$$

$$\rho_p = \frac{\gamma_1}{\gamma_0}$$

$$(\alpha_1 X_{t-1} + Z_t) X_{t-1}$$

Expectations:

$$\alpha_1 X_{t-1}^2 + Z_t X_{t-1}$$

$$\alpha_1 E[X_{t-1}^2] + E[Z_{t-1} * X_{t-1}] = \alpha_1 \gamma_0 + 0$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \alpha_1$$

$$\gamma_1 = \alpha_1 \gamma_0$$

[1] https://blackboard.le.ac.uk/bbcswebdav/pid-1477453-dt-content-rid-3910491_2/courses/MA4022/Time%20Series%201_2_3%282%29%281%29.pdf

References:

1)https://blackboard.le.ac.uk/bbcswebdav/pid-1477453-dt-content-rid-3910491_2/courses/MA4022/Time%20Series%201_2_3%282%29%281%29.pdf

2)https://mcs.utm.utoronto.ca/~nosedal/sta457/ar1-and-ar2.pdf