

EE627A Midterm

Name: _____

Q1. An MA(2) process is defined by $X_t = \epsilon_t + a_1\epsilon_{t-1} + a_2\epsilon_{t-2}$. Determine

1. a The expectation $E[X_t]$.
1. b The variance $\text{var}[X_t]$
1. c The first-order autocorrelation $\rho(1)$
1. d What is the value for the 3rd order autocorrelation $\rho(3)$?

- Q2. (a) Find the autocorrelation function(ACF) of the time series $X_t = W_t + \frac{5}{2}W_{t-1} - \frac{3}{2}W_{t-2}$, where the Gaussian noise $W_t \sim \mathcal{N}(0, 1)$.
- (b) Find the autocorrelation function(ACF) of the time series $X_t = W_t - \frac{1}{6}W_{t-1} - \frac{1}{6}W_{t-2}$, where the Gaussian noise $W_t \sim \mathcal{N}(0, 9)$.

Q3. Suppose you are given the following classification task: predict the target $Y \in \{0, 1\}$, given two real valued features $X_1 \in \mathcal{R}$ and $X_2 \in \mathcal{R}$. After some training, you learn the following decision rule:

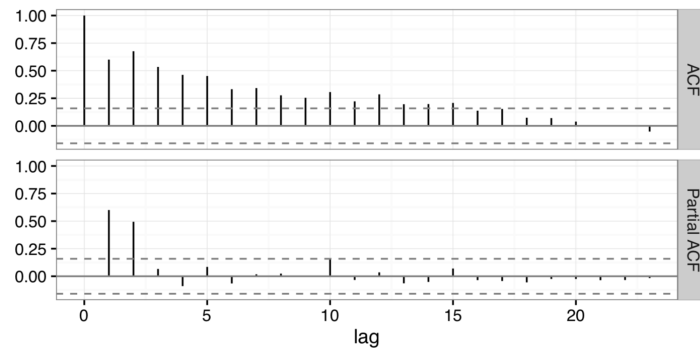
Predict $Y = 1$ if $w_1 X_1 + w_2 X_2 + w_0 \geq 0$ and $Y = 0$ otherwise, where $w_1 = 3$, $w_2 = 5$, and $w_0 = 15$.

Plot the decision boundary on the $\{x_1, x_2\}$ feature space and label the region where we would predict $Y = 1$ and $Y = 0$.

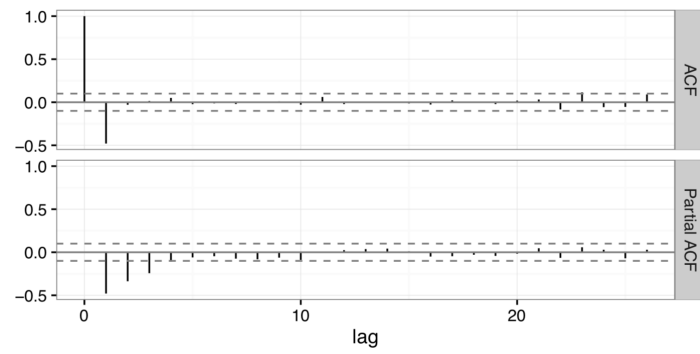
- Q4. We have two different methods to predict the college enrollments in a selected school. Given the predictive scores and real outcomes for each student, please plot the ROC curves for each method and compare the AUC values. Which predictive method is relatively better in terms of AUC value?

Table 1: Predictive Scores for Collenge Enrollments

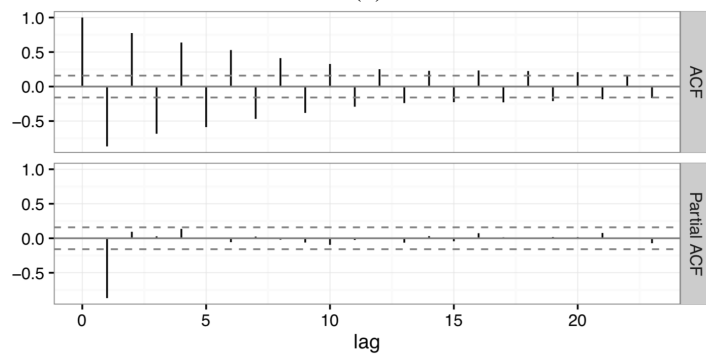
Student	Enrollment Decision	Method 1	Method 2
1	Yes	0.99	0.85
2	No	0.90	0.70
3	Yes	0.90	0.73
4	Yes	0.86	0.65
5	No	0.85	0.78
6	No	0.70	0.55



(a)



(b)



(c)

Q5. Above are the ACF and PACF for three time series. For each series, state whether it is autoregressive or moving average, and the order (p or q).

Q6. For a given predictor matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots]$ (say, students scores list), and the binary response vector \mathbf{y} (students enrolled to college or not), if we transform every column vector \mathbf{x}_i in \mathbf{X} to a normalized new vector

$$\mathbf{z}_i = \frac{\mathbf{x}_i - \text{mean}(\mathbf{x}_i)}{\text{standard deviation of } (\mathbf{x}_i)}$$

Does this normalization will change the AUC performance of logistic regression results? If yes, specify what the changes are? If not, explain why not?

