ALDA Fall 2018
HW 3
10/17/2018

HW3 contains 6 questions. Please read and follow the instructions.

- **DUE DATE FOR SUBMISSION: 11/8/2018 11: 45 PM**
- **TOTAL NUMBER OF POINTS: 100+5**
- Make sure you clearly list each team member's **name and Unity ID** at the top of your submission. One submission per group.
- Your submission should be a **single zip file** containing a PDF of your answers, your codes, and a readme file with running instructions. Please follow the naming convention for your zip file: H(homework group number)_HW(homework number), e.g. H1_HW3.

---

1. (12 points) [**D-Separation**][**Ruth Okoilu**]

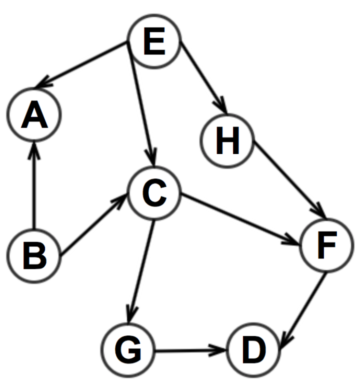
   Conditional independence is a key concept in Bayesian belief network. Please answer the following conditional independence and d-separation questions using the graphs below.

   (a) (3 points) In Figure 1 (left), are B and D d-separated given {A}? Justify your answer

   (b) (3 points) In Figure 1 (left), are A and D d-separated given {C, H}? Justify your answer.

   (c) (3 points) In Figure 1 (right), are A and B d-separated given {F, E}? Justify your answer

   (d) (3 points) In Figure 1 (right), are C and D d-separated given {B}? Justify your answer.



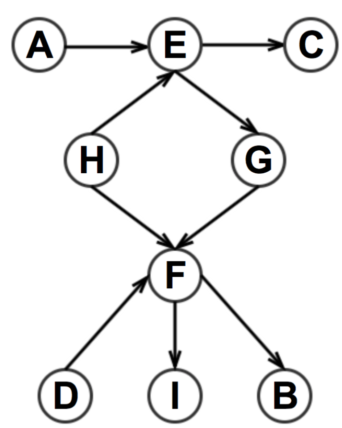


Figure 1: Two Bayesian belief networks

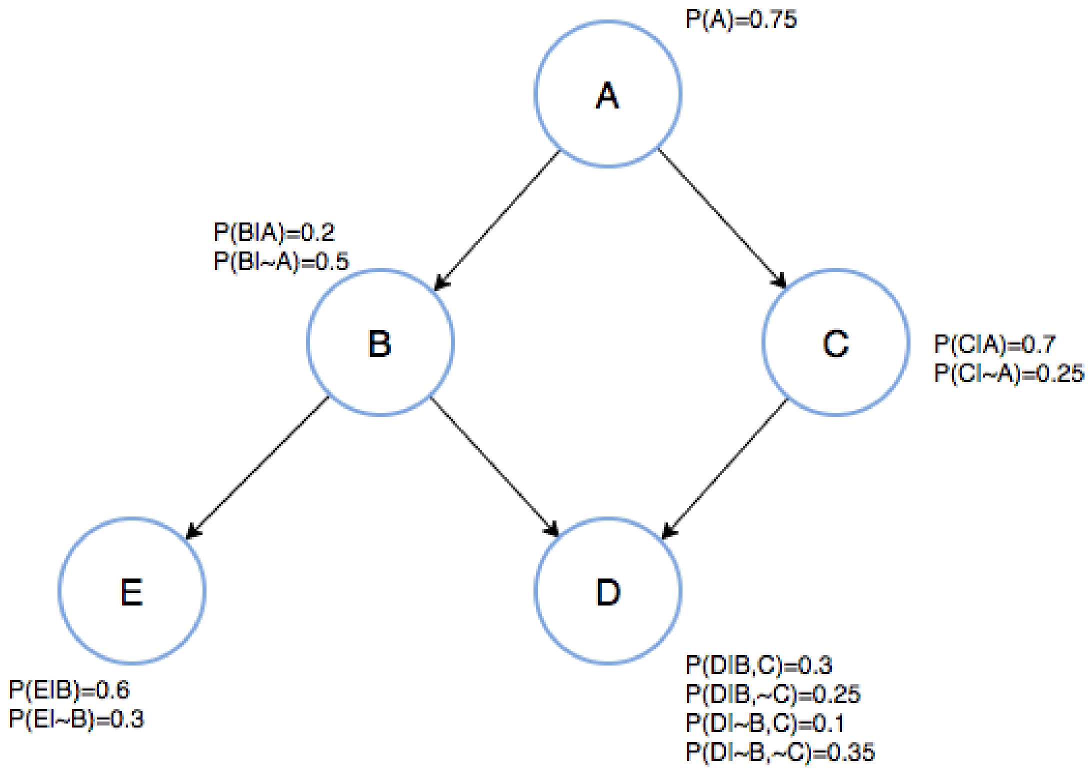


Figure 2: Q2: BN Inference

2. (15 points) [**BN Inference**][**Song Ju**] Compute the following probabilities according to the Bayesian net shown in Figure 2.
   (a) (5 points) Compute $P(E)$. Show your work.
   (b) (5 points) Compute $P(\sim B, C, D, E)$. Show your work.
   (c) (5 points) Compute $P(D \mid A)$. Show your work.

3. (20 points) [**LR**][**Xi Yang**]
   (a) (7 points) Given the following three data points of (x, y): $(1, 2)$, $(2, 1)$, $(0, -1)$, try to use a linear regression $y = \beta_1 x + \beta_0$ to predict y. Determine the values of β_1 and β_0 and show each step of your work.
   (b) (13 points) [Programming Task] Apply the following three linear regressions:
   (1) $y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_0$
   (2) $y = \beta_1 x_1^2 + \beta_2 x_2^2 + \beta_3 x_3^2 + \beta_4 x_4^2 + \beta_0$
   (3) $y = \gamma_1 x_1^3 + \gamma_2 x_2^3 + \gamma_3 x_3^3 + \gamma_4 x_4^3 + \gamma_0$
   to the provided data file "hw3q3(b).csv", which is from a combined cycle power plant dataset (`https://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant`). In the given data file, $x_i, i \in [1, 4]$ are four features and y is the prediction target which indicates hourly electrical energy output.

Write code in Matlab, R or Python to perform following tasks. Please *report your outputs and key codes* in the document file and also include your code (end with .m, .r or .py) in the .zip file.

i. (6 points) Load the data. Fit the whole dataset to the three linear regression models, respectively. *Report* the coefficients (αs, βs, γs) of the three models.

ii. (7 points) Use **leave-one-out** cross validation to determine the RMSE (root mean square error) for the three models. Specifically, in each fold, fit the training data to the model to determine the coefficients, then apply the coefficients to get predicted label for testing data (You don't need to report the coefficients in each fold). *Report* RMSE for the three models. Based on the RMSE, which model is the best for fitting the given data?

4. (16 points) (extra 5 points) [**ANN**] [**Ruth Okoilu**]

Train, validate, and test a neural network model using the dataset in hw3q4.zip, which contains training data (75%), validation data (12.5%), and test data (12.5%). There are two output classes in this data set. You can either choose matlab or a python neural networks package, Keras for this problem. (All the output should be included in your report. Otherwise, your points are deducted.)

(a) (5 extra points only for choosing Keras) Please briefly describe how to construct your working environments (e.g. language, package version, backend for neural networks, installation, etc.) in your report, and write how to execute your codes on 'readme' file.

(b) (8 points) (1) Construct neural networks using the given training dataset (X_train, Y_train) using different number of hidden neurons. Set the parameters as follows: activation function for hidden layer='relu', activation for output layer ='sigmoid', loss function ='mse', metrics= 'accuracy', epochs=10, batch_size=50. For each model, change the number of hidden neurons in the order of 2, 4, 6, 8, 10.
(2) Validate each neural network using the given validation dataset (X_val, Y_val). The validation accuracy is used to determine how many number of hidden neurons are optimal for this problem.
Provide the core code for "neural network learning" with comments in your report. (Please apply a fixed random seed 7 in order to generate a same result every time.)

(c) (3 points) Plot a figure, where the horizontal x-axis is the number of hidden neurons, and the vertical y-axis is the accuracy. Please plot both training and validation accuracy in your figure. (Note that the exact accuracy could be slightly different according to your working environments, however you can analyze the trend.)

(d) (3 points) Provide a simple analysis about your results and choose the optimal number of hidden neuron from the analysis.

(e) (2 points) Report the test accuracy using the given test dataset (X_test, Y_test) on the neural network with the optimal number of hidden neurons.

5. (22 points) [**SVM Theory**]

(a) (10 points) [**Song Ju**] Support vector machines (SVM) learn a decision boundary leading to the largest margin between classes. In this question, you'll train a SVM on a tiny dataset with 4 data points as shown in Figure 3. This dataset consists of two points with class1 (label 1) and two points with class2 (label -1).

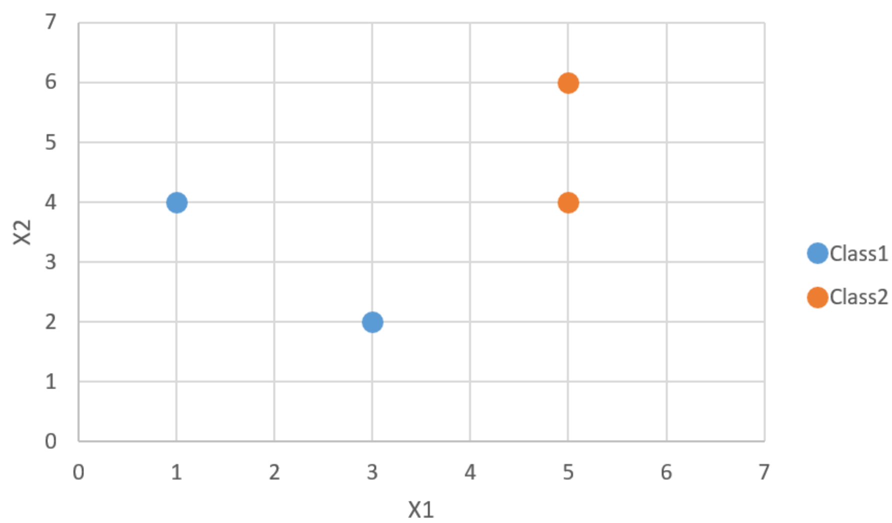


Figure 3: Q5(a)

i. (5 points) Find the weight vector $\mathbf{w}$ and bias b. What is the equation corresponding to the decision boundary?

ii. (5 points) Circle the support vectors and draw the decision boundary.

(b) (12 points) [**Xi Yang**] Given 2-dimensional data points $\mathbf{X}^i, i \in [1,2,3,4]$ as shown in Table 1, in this question, you will employ the kernel function for SVM to classify these four data points.

Data ID	x_1	x_2	y
$\mathbf{X}^1$	0	2	-1
$\mathbf{X}^2$	2	0	-1
$\mathbf{X}^3$	0	0	1
$\mathbf{X}^4$	2	2	1

Table 1: Q5(b)

i. (4 points) Suppose the kernel function is: $\mathbf{K}(\mathbf{X}^i, \mathbf{X}^j) = (1 + \mathbf{X}^i \cdot \mathbf{X}^j)^2$, where $\mathbf{X}^i$ and $\mathbf{X}^j$ indicate two data points. This kernel is equal to an inner product $\phi(\mathbf{X}^i) \cdot \phi(\mathbf{X}^j)$ with a certain function of ϕ. What is the function of ϕ?

ii. (2 points) Transform the four given data points $\mathbf{X}^i, i \in [1,2,3,4]$ to the higher dimensional space via the function ϕ get from (i). Report your results.

iii. (6 points) Assume the four transformed data points get from (ii) are all support vectors. Apply Lagrange multipliers to determine the *maximum margin linear decision boundary* in the transformed higher dimensional space.

6. (15 points) [**SVM Programming**][**Xi Yang**]

In this question, you will employ SVM to solve a classification problem for the provided data file "hw3q6.csv". Each row in the data file indicates a sample. The first 12 columns are features and the last column "Class" indicates the label, with 1 and 0 indicating the positive and negative samples, respectively.

Write code in Matlab, R or Python to perform the following tasks. Please *report your outputs and key codes* in the document file and also include your code (end with .m, .r or .py) in the .zip file.

(a) (1 point) Load data. *Report* the size of positive and negative samples in dataset.

(b) (4 points) Use *stratified random sampling* to divide the dataset into training data (75%) and testing data (25%). *Report* the number of positive and negative samples in both training and testing data.

(c) (4 points) Take SVM with linear kernel as classifier (third-party packages are allowed to use) and set the regularization parameter C as: [0.1, 0.5, 1, 5, 10, 50, 100], respectively. For each value of C, train a SVM classifier with the training data and get the number of support vectors (SVs). Generate a plot with C as the horizontal axis and number of SVs as the vertical axis. Give a brief analysis for the plot.

(d) (6 points) Compare 4 different kernel functions, including linear, polynomial, radial basic function (Gaussian kernel), and sigmoid kernel. Make a table to record the *accuracy*, *precision*, *recall* and *f-measure* of the classification results for the 4 kernel functions. Try to tune the parameters via grid search and report your best results with the optimal parameters. Based on the results, which kernel function will you choose?