DDA3020: Homework 3

April 16, 2023

Homework due: **11:59pm, April 29, 2023**. This assignment accounts for 15/100 of the final score.

1 Written Problems (50 points)

1. (20 points) **Computational Graph(CG) and Backpropagation**
   Consider the following classification MLP with one hidden layer:

$$\begin{aligned}
\boldsymbol{x} &= \text{ input } \in \mathbb{R}^D \\
\boldsymbol{z} &= \mathbf{W}\boldsymbol{x} + \boldsymbol{b}_1 \in \mathbb{R}^K \\
\boldsymbol{h} &= \text{ReLU}(\boldsymbol{z}) \in \mathbb{R}^K \\
\boldsymbol{a} &= \mathbf{V}\boldsymbol{h} + \boldsymbol{b}_2 \in \mathbb{R}^C \\
\mathcal{L} &= \text{ CrossEntropy } (\boldsymbol{y}, \text{softmax}(\boldsymbol{a})) \in \mathbb{R}
\end{aligned}$$

   where $\boldsymbol{x} \in \mathbb{R}^D, \boldsymbol{b}_1 \in \mathbb{R}^K, \mathbf{W} \in \mathbb{R}^{K\times D}, \boldsymbol{b}_2 \in \mathbb{R}^C, \mathbf{V} \in \mathbb{R}^{C\times K}$, where D is the size of the input, $\boldsymbol{y}$ is the groundtruth, K is the number of hidden units, and C is the number of classes.
   hint : $\text{ReLU}(a) = \max(a, 0) = a\mathbb{I}(a > 0)$, where $\mathbb{I}(e)$ is the indicator function.

$$\mathbb{I}(e) = \begin{cases} 1 & \text{if } e \text{ is true} \\ 0 & \text{if } e \text{ is false} \end{cases}$$

   You can use the fact that $\text{ReLU}'(a) = H(a)$, where $H(a)$ is the Heaviside step function $H(a) = \mathbb{I}(a > 0)$.

   (a) Please plot the computational graph for the loss function $\mathcal{L}$, i.e. draw the computational graph of forward pass. (10 points)

   (b) Derive the backward pass, i.e. calculate parameter derivative of each layer, including $\nabla_{\mathbf{V}}\mathcal{L}$, $\nabla_{b_2}\mathcal{L}$, $\nabla_{\mathrm{W}}\mathcal{L}$, $\nabla_{b_1}\mathcal{L}$. (10 points)

2. (10 points) **CNN**
   Consider the convolutional network defined by the layers below. The input shape is $32 \times 32 \times 1$ and the output is 10 neurons. Consider layers, $Conv5(10) + Maxpool_2 + Conv5(10) + Maxpool_2 + FC10$, where

- $Conv5(10)$: 10 filters with each size $5 \times 5 \times D$, where D is the depth of the activation volume at the previous layer, stride=1, padding=2
- $Maxpool_2$: 2×2 filter, stride=2, padding=0
- $FC10$: A fully-connected layer with 10 output neurons.

(a) Compute the shape of activation map of each layer. (5 points)

(b) Compute the total number of parameters of each layer. (5 points)

3. (10 points) **Decision Tree**
Consider the following table which predicts whether a student is likely to be overweight. Use Cross Entropy as criteria to construct the classification decision tree that predicts whether or not a student is likely to be overweight. Show your step of the computation.

Student ID	gender	Hyperlipidemia	unhealthy diet	exercises	overweight
1	boy	yes	no	yes	yes
2	boy	yes	yes	no	yes
3	girl	no	yes	no	yes
4	boy	no	no	yes	no
5	girl	yes	yes	yes	yes
6	boy	no	yes	yes	no

4. (10 points) Assessing AUC and Performance of a Binary Classifier

Given a 2d-dataset of 5 positive and 5 negative samples, and a simple linear model with a sigmoid function ($f(x)$), please complete the following tasks:

$$\text{Positive Samples: } \begin{bmatrix} (5 & 5) \\ (3 & 8) \\ (-1 & 8) \\ (5 & 4) \\ (-1 & -2) \end{bmatrix} \quad \text{Negative Samples: } \begin{bmatrix} (-5 & 1) \\ (2 & -2) \\ (-1 & 1) \\ (-5 & -10) \\ (-10 & -9) \end{bmatrix}$$

$$f(x) = \sigma(w_1 x_1 + w_2 x_2)$$

where σ is a sigmoid function, $w_1 = 0.5$, and $w_2 = -0.1$.

(a) Compute the predictions for all data points in the dataset using the provided model ($f(x)$) (2 points).

(b) Assume the threshold is 0.5, evaluate the performance of the model by calculating the following evaluation scores (3 points):

- Accuracy
- Precision

- Recall
- F1 Score
- Confusion Matrix

(c) Construct a ROC curve for the model as follows (3 points).

- The list of thresholds is $[0, 0.2, 0.4, 0.6, 0.8, 1.0]$ (use the step-wise curve, like Fig. 1).
- Hand-writing and Generating with software (e.g. Python) are both ok.
- x-axis is FPR and y-axis is TPR.
- Use the following graph as a reference.

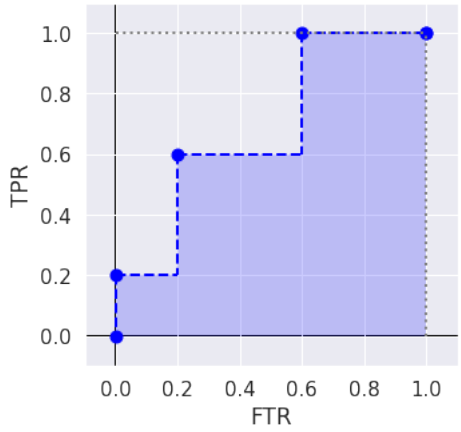

Figure 1: Example of step-wise ROC curve

(d) Calculate the Area Under the ROC Curve, i.e., AUC (2 points).

2 Programming (50 points)

1. Decision Tree (required) (25 points)

   Task description Classify the variable *species* in the **Penguins** dataset, using decision tree, bagging, and random forests. All these algorithms can be implemented by calling **sklearn** in Python, including "DecisionTreeClassifier", "BaggingClassifier" and "RandomForestClassifier".

   Dataset **Penguins** contains 345 data points (saved in 345 rows). For each data, the first column is the species of penguins that we want to classify; the remaining 6 columns indicate 6 features (or attributes). There is no fixed train/test splitting. In this project, you can randomly split the whole dataset with the ratio 75% for training set and 25% for test set, and try multiple times. The dataset can be obtained from `https://github.com/mwaskom/seaborn-data/blob/master/penguins.csv`

   What you should do

   - **Data preprocessing**: use **pandas** library to check the data in the dataset. Process incomplete data point such as 'NaN' or 'Null'. (1.5 points)
   - **Data statistics**: analyze the statistics of the target variable and each feature, and try to visualize the statistics (*e.g.*, histogram) (1.5 points)
   - **Decision tree**: solve the above problem using decision tree method; report the train/test accuracy with respect to three different maximum depths, three different least node sizes; plot the learned tree (6 points)
   - **Bagging of trees**: solve the above problem using the bagging method, with decision tree as the base learner; report the train/test accuracy with respect to three different depths, three different number of trees (6 points)
   - **Random forests**: solve the above problem using the random forest method, with decision tree as the base learner; report the train/test accuracy with respect to three different number of trees, three different values of m (the number of candidate attributes to split in every step, see Slides 'Decision Tree', Page 69) (6 points)
   - Plot the curve of bias^2 with respect to different number of trees in random forests, *e.g.*, $\#tree = 10, 20, \ldots, 100$. Then, describe the relationship between bias^2 and different number of trees; repeat the procedure for variance. You can use the function from `https://rasbt.github.io/mlxtend/user_guide/evaluate/bias_variance_decomp/` to get the results. (4 points)

2. Fashion-MNIST Recognition using sk-learn (25 points)

About Dataset Fashion-MNIST is a dataset of Zalando's article images—consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 grayscale image, associated with a label from 10 classes. The original graphics have been flattened into vector formula (784,). For more details, you can check the data files.

Task description Use the function "MLPClassifier" of sk-learn to construct a fully connected network to classify the Fashion-MNIST data. The dataset has been provided on BB as *train.csv* and *test.csv*. The training-testing splitting has been given. Build neural networks and print out the accuracy.

Experiment settings Show the performance of your neural network with different structures:

- number of hidden layers chosen from $\{1, 2, 3\}$
- number of hidden nodes chosen from $\{50, 200, 784\}$ (let all hidden layers have the same number of nodes)
- optimizers chosen from $\{Adam, SGD\}$
- other hyperparameters such as learning rate can be chosen as you wish. Just have a try!

Report Contents In the report, the following contents should be included.

- The meaning of hyperparameters and parameters in your code, for example, solver='adam' means that we use Adam optimizer to finish the gradient descent process.
- The prediction accuracy of each networks. Analysis the influence of parameters settings.

Note that you should submit A3_StudentID.pdf (report, together with the written answers), A3_StudentID_Tree.ipynb, and A3_StudentID_NN.ipynb (code files for Decision Tree and Fashion-MNIST). Do not zip them.
The reference report is in Assignment 1. You can check it on BlackBoard.
(You can submit several files in one submission. Don't submit them in different submissions.) Your report for the programming questions should include necessary formulas, charts, and explanations. The number of pages is suggested to be 2-5.

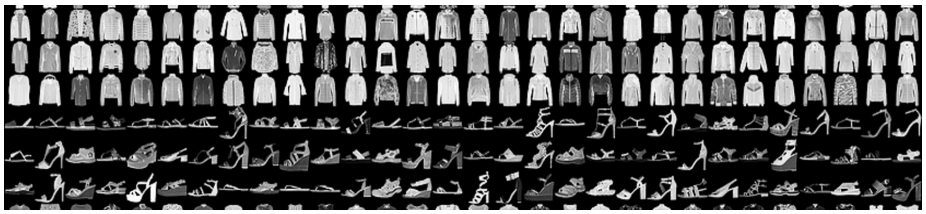

Figure 2: Fashion-MNIST Dataset