

Machine Learning – Hands-on Assignments

Instructions:

1. *Refer to the lesson names and tutorial reference in the questions given below.*
2. *Open and use the corresponding dataset from the dataset folder to find the solution.*

Lesson 2: Techniques of Machine Learning

Tutorial Reference: Predict whether consumers will buy houses or not, given their age and salary.

Q.1) What issues do you see in the plot produced by the code in reference to the above problem statement?

Tutorial Reference: Predict house prices based on the area of the house.

Q.2) What are the approximate prices of the houses with areas 1700 and 1900?

Lesson 3: Data Preprocessing

Tutorial Reference: Demonstrate methods to handle missing data, categorical data, and data standardization

Q.3) Review the training dataset (Excel file). Note that weight is missing for the fifth and eighth rows. What are the values computed by the imputer for these two missing rows?

Q.4) In the tutorial code, find the call to the Imputer class. Replace strategy parameter from “mean” to “median” and execute it again. What is the new value assigned to the blank fields Weight and Height for the two rows?

Q.5) In the code snippet given below in the tutorial, why does the array X have 5 columns instead of 3 columns as before?

Tutorial Reference: Demonstrate how to reduce data dimensions from 3D to 2D.

Q.6) What does the hyperplane shadow represent in the PCA output chart on random data?

Q.7) What is the reconstruction error after PCA transformation? Give interpretation.

Lesson 5: Regression

Tutorial Reference: Demonstrate how to reduce data dimensions from 3D to 2D.

Q.8) Modify the degree of polynomial from Polynomial Features (degree = 1) to 1, 2, 3, and interpret the resulting regression plot. Specify if it is underfitted, right-fitted, or overfitted.

Q.9) Predict the insurance claims for age 70 with polynomial regression n with degree 2 and linear regression.

Tutorial Reference: [Predict insurance premium per year based on a person's age using Decision Trees.](#)

Q.10) Modify the code to predict insurance claim values for anyone above the age of 55 in the given dataset.

Tutorial Reference: [Generate random quadratic data and demonstrate Decision Tree regression.](#)

Q.11) Modify the `max_depth` from 2 to 3 or 4, and observe the output.

Q.12) Modify the `max_depth` to 20, and observe the output.

Q.13) What is the class prediction for `petal_length = 3` cm and `petal_width = 1` cm for the `max_depth = 2`?

Q.14) Explain the Decision Tree regression graphs produced when `max_depths` are 2 and 3. How many leaf nodes exist in the two cases? What does average value represent for these two situations?

Q.15) Modify the regularization parameter `min_sample_leaf` from 10 to 6, and check the output of Decision Tree regression. What is the result and why?

Tutorial Reference: [Predict insurance per year based on a person's age using Random Forests.](#)

Q.16) What is the output insurance value for individuals aged 60 and with `n_estimators = 10`?

Tutorial Reference: [Demonstrate various regression techniques over a random dataset.](#)

Q.17) The program depicts a learning process when the values of the learning rate η are 0.02, 0.1, and 0.5. Give your interpretation of these charts?

Q.18) The program depicts the learning process when the values of the learning rate η are 0.02, 0.1, and 0.5. Try changing the values to 0.001, 0.25, and 0.9 and check the results? Provide interpretation.

Lesson 6: Classification

Tutorial Reference: [Predict if the consumers will buy houses, given their age and salary.](#)

Q.19) Typically, the value of `nearest_neighbors` for testing class in KNN is 5. Modify the code to change the value of `nearest_neighbours` to 2 and 20, and note the observations.

Tutorial Reference: Classify IRIS dataset using SVM, and demonstrate how Kernel SVMs can help classify non-linear data.

Q.20) Modify the kernel trick from RBF to linear to see the type of classifier that is produced for the XOR data in this program. Interpret the data.

Q.21) For the Iris dataset, add a new code at the end of this program to produce classification for RBF kernel trick with $\gamma = 1.0$. Explain the output.

Tutorial Reference: Classify IRIS flower dataset using Decision Trees

Q.22) Run decision tree on the IRIS dataset with max depths of 3 and 4, and show the tree output.

Q.23) Predict and print class probability for Iris flower instance with `petal_len` 1 cm and `petal_width` 0.5 cm.

Tutorial Reference: Classify the IRIS flower dataset using various classification algorithms.

Q.24) Add Logistic Regression classification to the program and compare classification output to previous algorithms?

Lesson 7: Unsupervised Learning with Clustering

Tutorial Reference: Demonstrate Clustering algorithm and the Elbow method on a random dataset.

Q.25) Modify the number of clusters `k` to 2, and note the observations.

Q.26) Modify the `n_samples` from 150 to 15000 and the number of centers to 4 with `n_clusters` as 3. Check the output, and note your observations.

Q.27) Modify the code to change the `n_samples` from 150 to 15000 and number of centers to 4, keeping `n_clusters` at 4. Check the output.

Q.28) Modify the number of clusters `k` to 6, and note the observations.