# Overview

| Question | Points | |
|---|---|---|
| Key Concepts of the Newsvendor Problem | 10 | |
| Contextual Features, Linear Regression, and Time Series Models | 10 | |
| Deep Learning and Neural Networks in the Newsvendor Problem | 10 | |
| Decision Trees and Random Forests | 10 | |
| Configuring the code of a Deep Learning model | 6 | |
| Linear Regression Newsvendor | 6 | |
| Decision Tree and Random Forest | 8 | |
| Total: | 60 | |

Question 1 Key Concepts of the Newsvendor Problem ..... *10 Points*

(a) (2 Points) What is the primary goal of the Newsvendor Problem" discussed in the session?

A. To minimize the total number of orders

**B. To balance underage and overage costs**

C. To predict future market trends

D. To increase customer satisfaction with faster delivery times

(b) (2 Points) In the ZenBite case example, what is the purpose of splitting the data into a training and test dataset?

A. To ensure that both datasets have the same demand values

**B. To evaluate the model's performance on unseen data**

C. To ensure a random distribution of data

D. To separate high-demand and low-demand products

(c) (2 Points) What is the main advantage of using the average demand for planning in the ZenBite case study?

**A. It provides a simple, easy-to-implement baseline strategy**

B. It guarantees zero underage and overage costs

C. It ensures the lowest possible cost for every product

D. It adjusts dynamically to fluctuations in demand

(d) (2 Points) What is the main idea behind the Sample Average Approximation (SAA) method?

A. To calculate the optimal order quantity based on a normal distribution of demand

B. To maximize the fulfillment of daily demand using the highest observed demand in the data

C. To defrost the same quantity each day based on the median demand of past data

**D. To minimize average costs by estimating the optimal order quantity based on historical demand data**

(e) (2 Points) Briefly explain how to calculate the optimal service level $\alpha$ and what it represents?

**Solution:** $\alpha$ is the ratio of underage and overage costs, it represents how often we want to fulfill the total demand

$$\alpha = \frac{cu}{cu + co}$$

Question 2 Contextual Features, Linear Regression, and Time Series Models . . . . *10 Points*

(a) (2 Points) What is a contextual feature in the Newsvendor Problem?

**A. Additional information available at the time of decision-making that may influence demand**

B. A feature used to directly calculate the demand based on past values

C. A specific type of data that includes only numeric variables like sales figures or temperature

D. Historical data that is only used to validate the accuracy of the model

(b) (2 Points) How does the linear regression Newsvendor model differ from ordinary linear regression?

A. Linear regression Newsvendor minimizes the sum of squared errors between actual and predicted demand

**B. Linear regression Newsvendor uses a loss function that penalizes underordering different from overordering**

C. Linear regression Newsvendor only applies to time series data

D. There is no significant difference between the two

(c) (2 Points) What is overfitting in the context of machine learning?

A. Fitting the model to only the first half of the dataset

**B. When a model performs well on training data but poorly on test data due to being too complex**

C. Underestimating future demand consistently

D. Using too few contextual features in the model

(d) (2 Points) Briefly explain what autocorrelation in a time series means?

**Solution:** Autocorrelation in a time series measures how current values are related to past values at different time lags.

(e) (2 Points) What is an example for incorporating seasonality into a time series model?

A. By adding a lag feature for the last day

**B. By including the month of the year as a feature**

C. By ignoring trend components in the data

D. By ensuring the model only considers weekdays

Question 3 Deep Learning and Neural Networks in the Newsvendor Problem . . . . *10 Points*

(a) (2 Points) What is the primary purpose of using deep learning in the Newsvendor Problem?

A. To replace all previous models with a more complex algorithm
B. To reduce the number of features needed for accurate demand prediction
**C. To account for non-linear relationships between features and demand, and capture interactions between features**
D. To eliminate the need for training data

(b) (2 Points) Why do we train deep learning models in an iterative process over multiple epochs?

**A. To gradually adjust the model's parameters and improve its predictions with each pass through the data**
B. To ensure the model reaches a global minimum of the loss function on the first pass
C. To avoid using too much data at once, which could cause the model to overfit
D. To ensure that the model changes weights randomly with each iteration

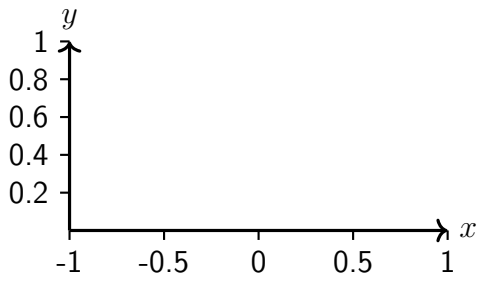
(c) (2 Points) Why is the learning rate important when training deep learning models?

**A. It controls how much the model's parameters are adjusted at each step during training, impacting how quickly or slowly the model learns**
B. It determines how many layers the model will have
C. It ensures that the training data is normalized before being used in the model
D. It prevents the model from overfitting by limiting the number of parameters learned

(d) (2 Points) What is the role of hyperparameters in training a deep learning model?

**A. Hyperparameters control the structure and training process of the neural network, such as the number of layers, learning rate, and batch size**
B. Hyperparameters are automatically optimized by the model and do not require manual tuning
C. Hyperparameters determine the weights of the connections between neurons
D. Hyperparameters are used to adjust the features being input into the model

(e) (2 Points) Draw the ReLU activation function in the coordinate system below.


y
1
0.8
0.6
0.4
0.2
x
-1
-0.5
0
0.5
1

Question 4 Decision Trees and Random Forests ..... *10 Points*

(a) (2 Points) What is the role of a Decision Tree when solving the Newsvendor Problem?

A. To provide a single numerical prediction for demand based on all features combined

B. To improve the accuracy of linear models by reducing error rates

C. To avoid the need for optimization steps by providing pre-calculated solutions

**D. To split data into smaller subsets based on the most informative features**

(b) (2 Points) In the context of Decision Trees, what are leaf nodes?

**A. The final nodes in the tree that contain data points grouped by similar characteristics, used to make predictions**

B. The initial nodes that contain all the data before any splits are made

C. Intermediate nodes that contain predictions for each feature

D. Nodes that perform optimization on the predicted data

(c) (2 Points) Why are Random Forests more robust than individual Decision Trees?

A. Random Forests use deep learning techniques to enhance model performance

B. Random Forests are trained on more features than individual decision trees

**C. Random Forests combine multiple trees, to reduce variance and improve accuracy of the predictions**

D. Random forests eliminate the need for feature engineering

(d) (2 Points) How does the use of multiple decision trees in a random forest help avoid overfitting?

**A. Each tree is trained on a random subset of the data, reducing the reliance on specific patterns and improving generalization**

B. Random forests apply pruning techniques to limit the depth of each tree

C. Random forests automatically drop irrelevant features during training

D. The use of a single strong learner in random forests helps minimize overfitting

(e) (2 Points) What does it imply if your trained Decicion Tree has exactly one observation in each leaf node?

**Solution:** It implies that the model overfitted.

Question 5 Configuring the code of a Deep Learning model ..... *6 Points*

Below is a snippet of Python code used to train a Deep Learning Newsvendor model:

```
1  dlnv = DeepLearningNewsvendor(
2                  cu=15,
3                  co=5,
4                  neurons=[25, 12],
5                  activations=['ReLU', 'ReLU'],
6                  optimizer='adam',
7                  epochs=250,
8                  batch_size=64,
9                  learning_rate=0.003,
10                 drop_prob=0.15,
11                 random_state=None,
12                 verbose=1,
13                 print_freq=10
14 )
15
16 dlnv.fit(
17         X_train,
18         demand.to_numpy()[train],
19         validation_data = (X_val, demand.to_numpy()[val]))
20
21 q_dlnv = dlnv.predict(X_val)
22
23 avg_cost_dlnv = average_costs(demand.to_numpy()[val], q_dlnv, cu=cu, co=co)
24
25 print(f"The final validation performance is: {avg_cost_dlnv:.2f}\)
```

(a) (2 Points) What do you have to change to train the model to have a higher service level than 75%?

**Solution:** Increase $cu$ or decrease $co$

(b) (2 Points) What do you have to change in the code to use a three-layer network with 16 neurons in each layer.

**Solution:** Change the code in row 4 to `neurons=[16,16,16]` and change the code in row 5 to `activations=['ReLU', 'ReLU', 'ReLU']`

(c) (2 Points) You suspect that your model overfits. Name two changes in the code above that would help to react to this problem.

**Solution:** Increase `drop_prob` or decrease the number of `neurons`

Question 6 Linear Regression Newsvendor ..... *6 Points*

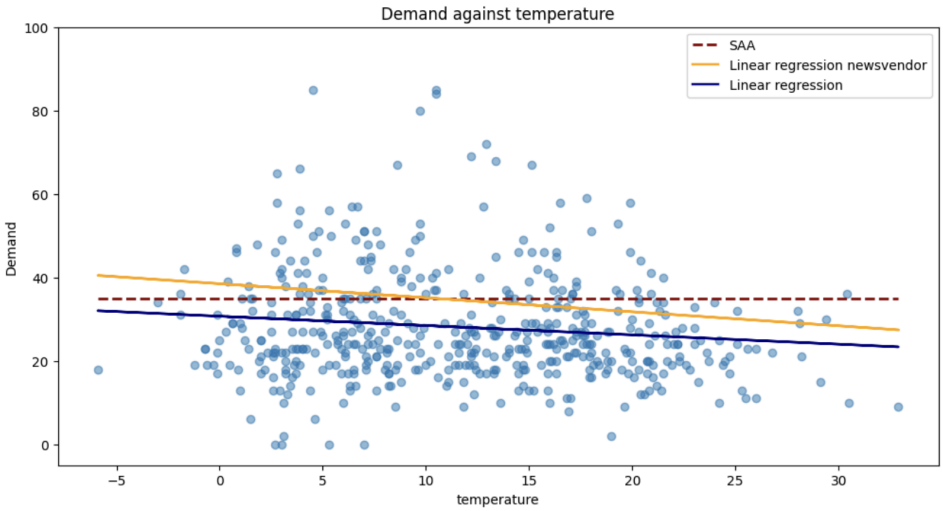
In the following figure you see the output of the models trained in the code below.

```
from ddop2.newsvendor import LinearRegressionNewsvendor
from sklearn.linear_model import LinearRegression
from ddop2.newsvendor import SampleAverageApproximationNewsvendor

# Define the cost structure
cu = 15
co = 5

# Create the models
lr = LinearRegression()
lrnw = LinearRegressionNewsvendor(cu, co)
SAA=SampleAverageApproximationNewsvendor(cu, co)

# Fit the models
lr.fit(features.loc[train][['temperature']], demand[train])
lrnw.fit(features.loc[train][['temperature']], demand[train])
SAA.fit(demand[train])
```




Please answer the following questions:

(a) (2 Points) How does the red line (SAA) change if we increase $cu$?

**Solution:** Moves up because it is more expensive to understock

(b) (2 Points) How does the blue line (Linear Regression) change if we increase $co$?

**Solution:** Unchanged because Linear Regression minimizes MSE

(c) (2 Points) How does the yellow line (Linear Regression Newsvendor) change if we decrease $cu$?

**Solution:** Moves down because the costs of overstocking decrease

Question 7 Decision Tree and Random Forest ..... *8 Points*

(a) (6 Points) In the table below you find the results of a prediction using a Random Forest for the next day. Use the output to calculate the optimal order quantity given underage costs ($cu$) of $15$ and overage costs ($co$) of 5.

| **Sample** | **Weekend** | **Temperature [in °C]** | **Demand** | **# Occurrences in leaf nodes across trees** | **Weight** |
|---|---|---|---|---|---|
| 1 | 0 | 19 | 27 | 2 | 2/10 |
| 5 | 0 | 24 | 20 | 5 | 5/10 |
| 6 | 0 | 22 | 23 | 3 | 3/10 |
| **Total:** | | | | 10 | 1 |

Tabelle 1: Sample Table with Occurrences and Weight

**Solution:** Step 1 - Calculate service level:

$$\alpha = \frac{15}{15+5} = 75\% \tag{1}$$

Step 2 - Find the 0.75 quantile:

$$d^{ordered} = [20, 23, 27] \tag{2}$$

$$(1 - p^{weighted}) = [0.5, 0.8, 1] \tag{3}$$

We get an optimal order quantity of 23

(b) (2 Points) We used the following code to train the model:

```
from ddop2.newsvendor import RandomForestWeightedNewsvendor

# Set-up the model including the cost structure
rfnv = RandomForestWeightedNewsvendor(cu=cu,co=co,
                  criterion="squared_error",
                  n_estimators=100,
                  min_samples_split=5,
                  min_samples_leaf=3,
                  min_weight_fraction_leaf=0.,
                  max_features=0.9,
                  max_leaf_nodes=None,
                  )
```

What happens if we set `min_samples_leaf` to 5?

A. The new table will have exactly 5 rows
B. The new table will have at most 5 rows
**C. The new table will have a minimum of 5 rows**
D. The new table will have the same number of rows