/

Ds - 6.

Q.1. Why do we need to evaluate our models before model deployment?

→ - Model evaluation is an integral part of the model development process.

- It helps to find the best model that represents our data and how well the chosen model will work in the future.

- Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic & overfitted models.

- Evaluating a model is a core part of building an effective machine learning model.

- There are several evaluation metrics, like confusion matrix, cross-validation, AUC-ROC curve, etc.

- Different evaluation metrics are used for different kinds of problems.

- By the time we arrive at evaluation phase, modeling phase has already generated one or more candidate models.

- Critical importance that models are

evaluated before deployment.

- Deployment of models usually represents a capital expenditure & investment on the part of the company.

- If the models in question are invalid, then the company's time & money are wasted.

- We examine model evaluation techniques for each of the six main tasks of data mining:

① description ④ classification
② estimation ⑤ clustering
③ prediction ⑥ association.

Q.2. How is the square root of the Mean Square Error (MSE) interpreted?

→ - For estimation & prediction models, we are provided with both the estimated value $\hat{y}$ & the actual value y.

- Therefore, a natural measure to assess model adequacy is to examine the estimation error $(y - \hat{y})$.

- Usual measure used to evaluate estimation or prediction models is the mean square error (MSE)

$$MSE = \frac{\Sigma_i (y_i - \hat{y}_i)^2}{n - p - 1}$$

where p represents the number of model variables.

- Models that minimize MSE are preferred.

- Square root of MSE can be regarded as an estimate of typical error.

- This is known as standard error of estimate & denoted by $s = \sqrt{MSE}$

Q.3. What might be drawback of evaluation measures based on squared error? How might we avoid this?

→ - For estimation & prediction models, we are provided with both the estimated value $\hat{y}$ & actual value y.

- Therefore, a natural measure to assess model adequacy is to examine the estimation error ~~ty~~ $(y - \hat{y})$.

- Usual measure used to evaluate estimation or prediction models is the mean square error.

$$MSE = \frac{\Sigma_i (y_i - \hat{y}_i)^2}{n - p - 1}$$

where p represents number of model variables.

- square root of MSE can be regarded as an estimate of typical error.

- This is known as standard error of estimate & denoted by $s = \sqrt{MSE}$.

- An evaluation measure that was related to MSE is sum of squared errors (SSE)

$$SSE = \sum_{\text{records}} \sum_{\text{output nodes}} (\text{actual output})^2$$

- sum of squares total,

$$SST = \sum_{i=1}^{n} (y - \bar{y})^2$$

- sum of squares regression,

$$SSR = \sum_{i=1}^{n} (\hat{y} - \bar{y})^2$$

- One of the drawbacks of above evaluation measures is influence of outliers.

- This is because the above measures are based on the squared error, which is much larger for outliers than for bulk of data.

- Thus the analyst may prefer to use the Mean Absolute Error (MAE).

- $$MAE = \frac{\Sigma |y_i - \hat{y}_i|}{n}$$

- Calculating MAE.
① Calculate the estimated target values, $\hat{y}_i$.
② find the absolute value between each estimated value & its associated actual target value, y_i, giving you $|y_i - \hat{y}_i|$.
③ find the mean of absolute values from step ②. This is MAE.

Q. 4. Explain model evaluation techniques for the estimation & prediction tasks.
→ same as Q.3.

Q. 5. Explain model evaluation measures for the classification tasks.
→ We examine following evaluative concepts -
① model accuracy
② overall error rate
③ sensitivity & specificity
④ false-positive rate & false negative rate
⑤ Proportions of true positives & true negatives
⑥ Proportions of false positives & false negatives.
⑦ misclassifications costs & overall model cost.

(ii) error benefit table.
(iii) lift charts
(iv) gains charts.

(v) - Confusion matrix -

Let TN, FN, FP & TP represent the numbers of true negative, false negatives, false positives & true positives resp.

Let TAN = total actually negative = TN + FP
TAP = total actually positive = FN + TP
TPN = total predicted negative = TN + FN
TPP = total predicted positive = FP + TP.

N = TN + FN + FP + TP represent the grand total of counts in the four cells.

- Accuracy -

Accuracy represents an overall measure of proportion of correct classifications.

$$\text{accuracy} = \frac{TN + TP}{TN + FN + FP + TP} = \frac{TN + TP}{N}$$

- overall error rate -

Overall error rate measures proportion of incorrect classifications.

$$\text{overall } \cancel{\text{accuracy}} \text{ error rate} = 1 - \text{accuracy} = \frac{FN + FP}{N}$$

- sensitivity & specificity -
sensitivity measures the ability of the model to classify a record positively.

$$\text{sensitivity} = \frac{\text{no. of true positives}}{\text{total actually positive}} = \frac{TP}{TP + FN}$$

Specificity measures the ability to classify a record negatively.

$$\text{specificity} = \frac{\text{no. of true negatives}}{\text{total actually negatives}} = \frac{TN}{FP + TN}$$

- False positive rate & false negative rate -
These are additive inverses of sensitivity & specificity.

$$\text{false positive rate} = 1 - \text{specificity}$$
$$= \frac{FP}{FP + TN}$$

$$\text{false negative rate} = 1 - \text{sensitivity}$$
$$= \frac{FN}{TP + FN}$$

- Proportions of TP, TN, FP & FN.

$$\text{Proportion of } TP = PTP = \frac{TP}{TPP} = \frac{TP}{FP + TP}$$

$$\text{proportion of } TN = PTN = \frac{TN}{FN + TN}$$

proportions of FP = 1 − PTP = $\frac{FP}{FP + TP}$

proportions of FN = 1 − PTN = $\frac{FN}{FN + TN}$.

Proportion of false positive & proportion of false negative are additive inverses of proportion of true positive & proportion of true negative, resp.

- lift charts & gains charts.

For classification models, lift is a concept which seeks to compare response rates with & without using classification model. Lift charts & gain charts are graphical evaluative methods for assessing & comparing the usefulness of classification models.

$$\text{lift} = \frac{\text{proportion of true positives}}{\text{proportion of positive hits}} = \frac{TP/TPP}{TAP/N}.$$

can be used to compare model performance.

Q.6. Explain classification evaluation measures accuracy, overall error rate, sensitivity & specificity.

→ accuracy & overall error rate − same as in Q.5.

sensitivity & specificity -
same as Q.5.
- In some fields, sensitivity is referred to as recall.
- A perfect classification model would have sensitivity = 1.0 = 100%.
- A good classification model should have acceptable levels of both sensitivity & specificity.

Q.7. What is the difference between total predicted negative & total actually negative?
→ Q.5.

Q.8. What is relationship betⁿ accuracy & overall error rate?
→ Q.5.

Q.9. Explain classification evaluation measures false positive, false negative, proportions of true neg, true positive, false neg. & false positive.
→ Q.5.

Q.10. Explain how misclassification cost can be adjusted to reflect real world concerns.
→ - Consider the situation from the standpoint of the lending institution.
- If lender commits a false negative, an applicant who had high income gets turned down for a loan.

- If lender commits false ~~negative~~ positive, an applicant who had low income would be awarded loan.

- Hence false positive is considered more damaging from lender's point of view.

- Hence prefer to minimize proportion of false positives.

- Misclassification cost adjustment to affect the performance of algorithm -
① Proportion of false positives should decrease → since cost of making such an error is doubled.
② Proportion of false negatives should increase → because fewer false positives usually means more false negatives.
③ sensitivity should decrease.
④ specificity should increase.

Q.11. With suitable example explain decision cost / benefit analysis.

→ - Company managers may require that model comparisons be made in terms of cost / benefit analysis.
- For example, in comparing original model before cost adjustment (model 1) against model using cost adjustment (model 2).

- Managers may prefer to have resp. error rates, false negatives & false positives translated into dollars & cents.

- Analysts can provide model comparison in terms of anticipated profit or loss by associating a cost or benefit with each of four possible combinations of correct & incorrect classification.

outcome	classification	actual value	cost	rationale.
TN	≤ 50000	≤ 50000	\$0	no money gained / lost
TP	> 50000	> 50000	- \$300	anticipated avg interest revenue from loans.
FN	≤ 50000.	> 50000	\$0	no money gained / lost.
FP	> 50000	≤ 50000	\$500	cost of loan default avg over all loans to ≤ 50000 group.

cost of model 1 - false positive cost not doubled.

cost of model 2 - false positive cost doubled.
Negative costs represent profits.

Page No.	
Date	

Qs. 12. Explain use of lift charts & gains charts to compare model performance.

→ - For classification models, lift is a concept, which seeks to compare response rates with & without using classification model.

- Lift charts & gain charts are graphical evaluative methods for assessing & comparing the usefulness of classification models.

- A good classification model should identify in its positive classifications, a group that has higher proportion of positive hits than database as a whole.

- The concept of lift quantifies this.

- $$\text{lift} = \frac{\text{proportion of TP}}{\text{proportion of positive hits}} = \frac{TP/TPP}{TAP/N}$$

- Lift charts are often presented in their cumulative form, where they are denoted as cumulative lift charts or gain charts.

- Lift charts & gain charts can also be used to compare model performance.