Quiz 2 - Introduction to Unsupervised Learning Techniques

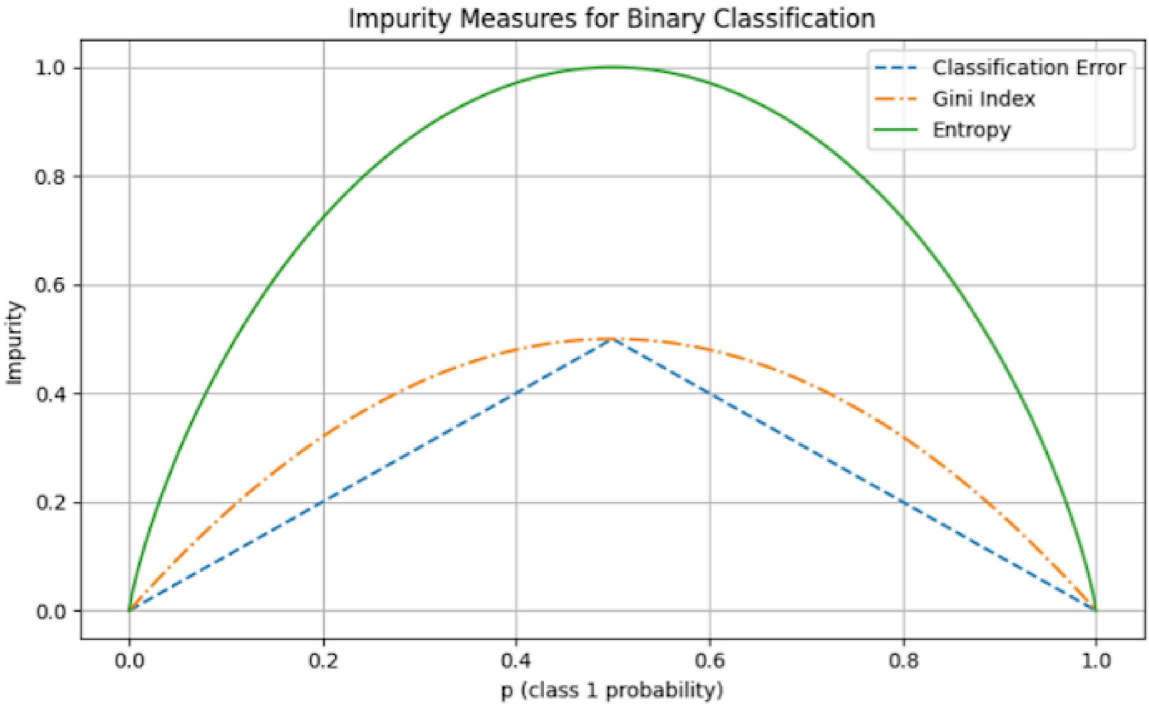
* Indicates required question

1. Name *

2. Email *

Decison Trees

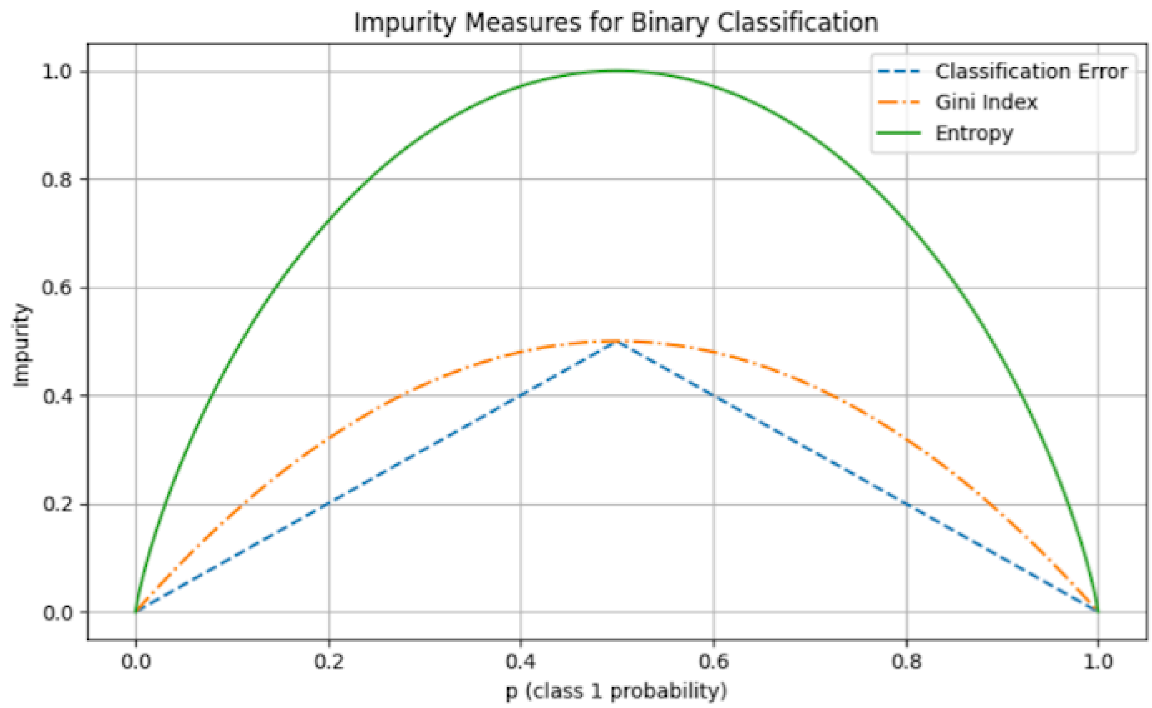
3. What is the entropy of a node where the label distribution is uniform (i.e., each class occurs with equal probability)? 1 point



Mark only one oval.

- Entropy is maximum
- Entropy is minimum
- Entropy is 0.5

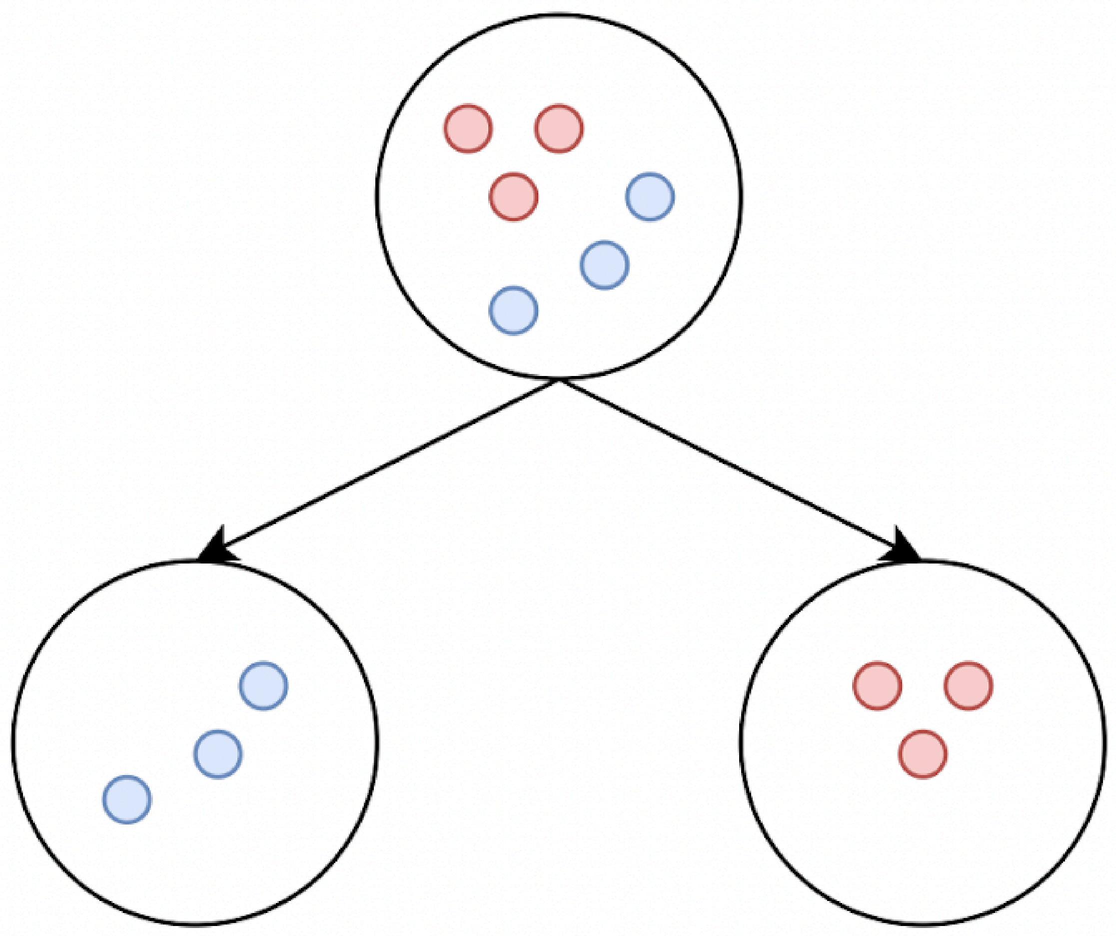
4. What is the entropy of a node that is pure (i.e., contains only one class)? 1 point



Mark only one oval.

- [ ] Entropy is maximum
- [ ] Entropy is minimum
- [ ] Entropy is 0.5

5. Based on the diagram, what is the Information Gain (IG) of this split? 1 point



Mark only one oval.

- ⬭ 0
- ⬭ 0.5
- ⬭ 1

Feature Importance Analysis

6. Which of the following is **NOT** a property of Mean Decrease Impurity (MDI)? 1 point

Algorithm Mean Decrease Impurity (MDI) for Random Forests

Require: Trained random forest model with T trees
Ensure: Feature importance scores $\{\mathrm{MDI}_j\}_{j=1}^{d}$

Initialize importance scores: $\mathrm{IMP}_j = 0$ for all features j
for each tree $t = 1$ to T in the forest **do**
    for each internal node n in tree t **do**
        Identify the feature F_j used for splitting at node n
        Let w_n be the proportion of samples reaching node n
        Calculate information gain $IG_n(j, \tau_n)$
        Update importance: $\mathrm{IMP}_j = \mathrm{IMP}_j + w_n \cdot IG_n(j, \tau_n)$
    end for
end for
Compute $\mathrm{MDI}_j = \frac{1}{T} \cdot \frac{\mathrm{IMP}_j}{\sum_{k=1}^{d} \mathrm{IMP}_k}$ for all j
return $\{\mathrm{MDI}_j\}_{j=1}^{d}$

Mark only one oval.

- It is an out-of-sample measure
- It is biased toward high-cardinality features
- It is specific to tree-based models

7. Which of the following is **NOT** a property of Permutation Feature Importance (PFI)? 1 point

Algorithm Permutation Feature Importance (PFI)

Require: Fitted model m, validation data D, repetitions K
Ensure: Feature importance scores $\{PFI_j\}_{j=1}^{d}$ with stds $\{\sigma_j\}_{j=1}^{d}$

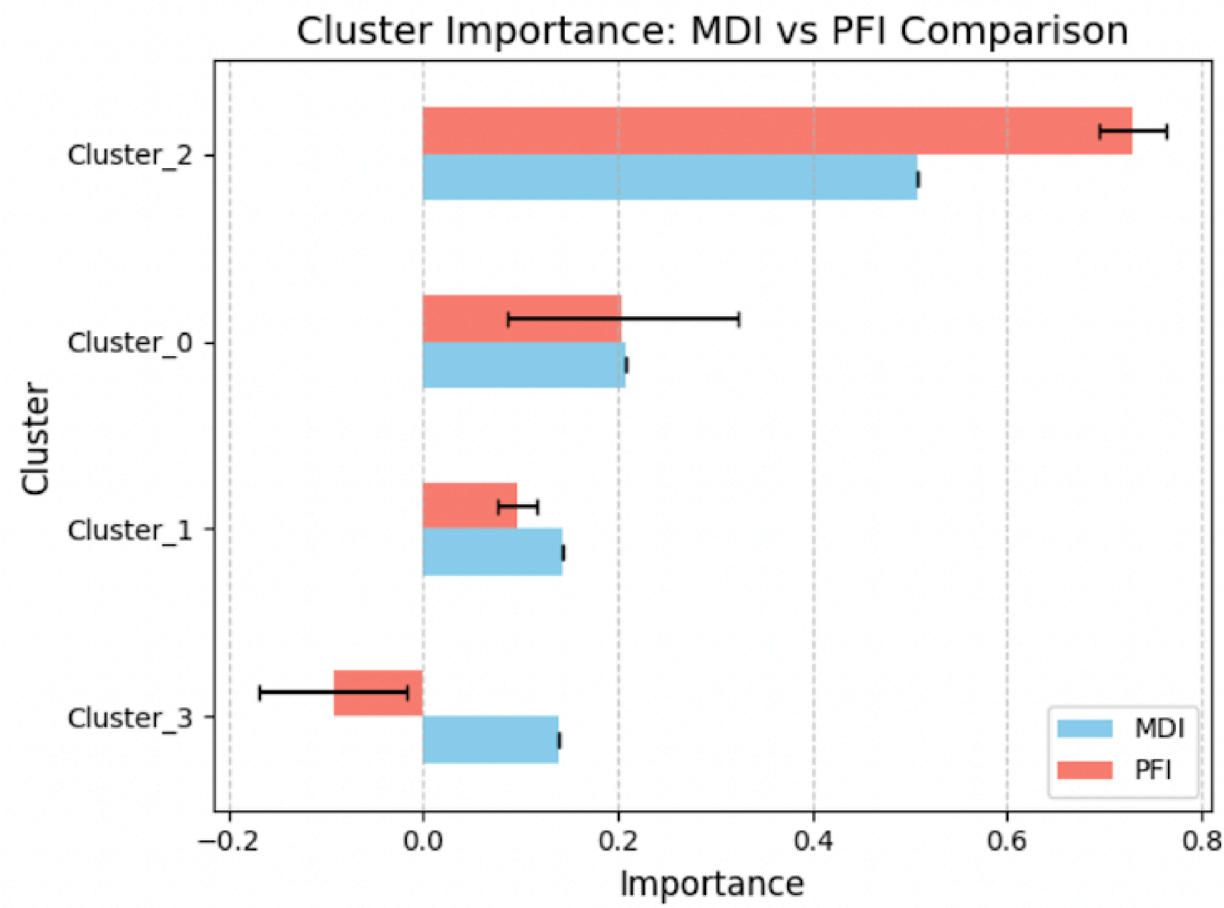
```
Compute reference score s of model m on data D
for each feature F_j (column of D) do
    Initialize array scores_j of length K
    for each repetition k in 1, ..., K do
        Randomly shuffle column j of dataset D to generate corrupted version D̃_{k,j}
        Compute score s_{k,j} of model m on corrupted data D̃_{k,j}
        Store in array: scores_j[k] = s − s_{k,j}
    end for
    Compute mean importance PFI_j and standard deviation σ_j from array scores_j
end for
return {PFI_j}_{j=1}^d and {σ_j}_{j=1}^d
```

Mark only one oval.

- ( ) It is agnostic to the evaluation metric
- ( ) It doesn't suffer from the substitution effect when features are correlated
- ( ) It is an out-of-sample measure

8. Based on the image from the programming session of feature importance at the cluster level, Cluster 3 has **positive MDI** and **negative PFI**. What can you deduce?

1 point



Mark only one oval.

- Cluster 3 contains noise features that the model mistakenly used during training
- Cluster 3 is highly informative, but PFI underestimates it due to randomness
- PFI is incorrect, because feature importance cannot be negative

K-means and GMMs

9. What does the K-means algorithm aim to minimize? 1 point

Algorithm The K-means Algorithm

Require: A data set $X = \{x_1, \dots, x_n\}$ $(x_i \in \mathbb{R}^p)$
Ensure: An assignment function Ψ^* and the associated centroids $c_1^*, \dots, c_K^*$.

Initialization: Choose $c_1, \dots, c_K$ in X at random
repeat
  for $i = 1 \dots n$ **do**
    $\Psi(x_i) \leftarrow \arg\min_{k \in \{1,\dots,K\}} \|x_i - c_k\|^2$
  end for
  for $k = 1 \dots K$ **do**
    $c_j \leftarrow \frac{1}{\sum_{i=1}^{n} \mathbb{1}(\Psi(x_i)=k)} \sum_{i=1}^{n} \mathbb{1}\left(\Psi(x_i) = k\right) x_i$
  end for
until convergence
return $\Psi^*, c_1^*, \dots, c_K^*$

Mark only one oval.

- ◯ The number of clusters
- ◯ The within-cluster distortion
- ◯ The average silhouette score

10. What is the expected silhouette score of a point that lies exactly between two clusters (equidistant to both)? 1 point

For each sample i:

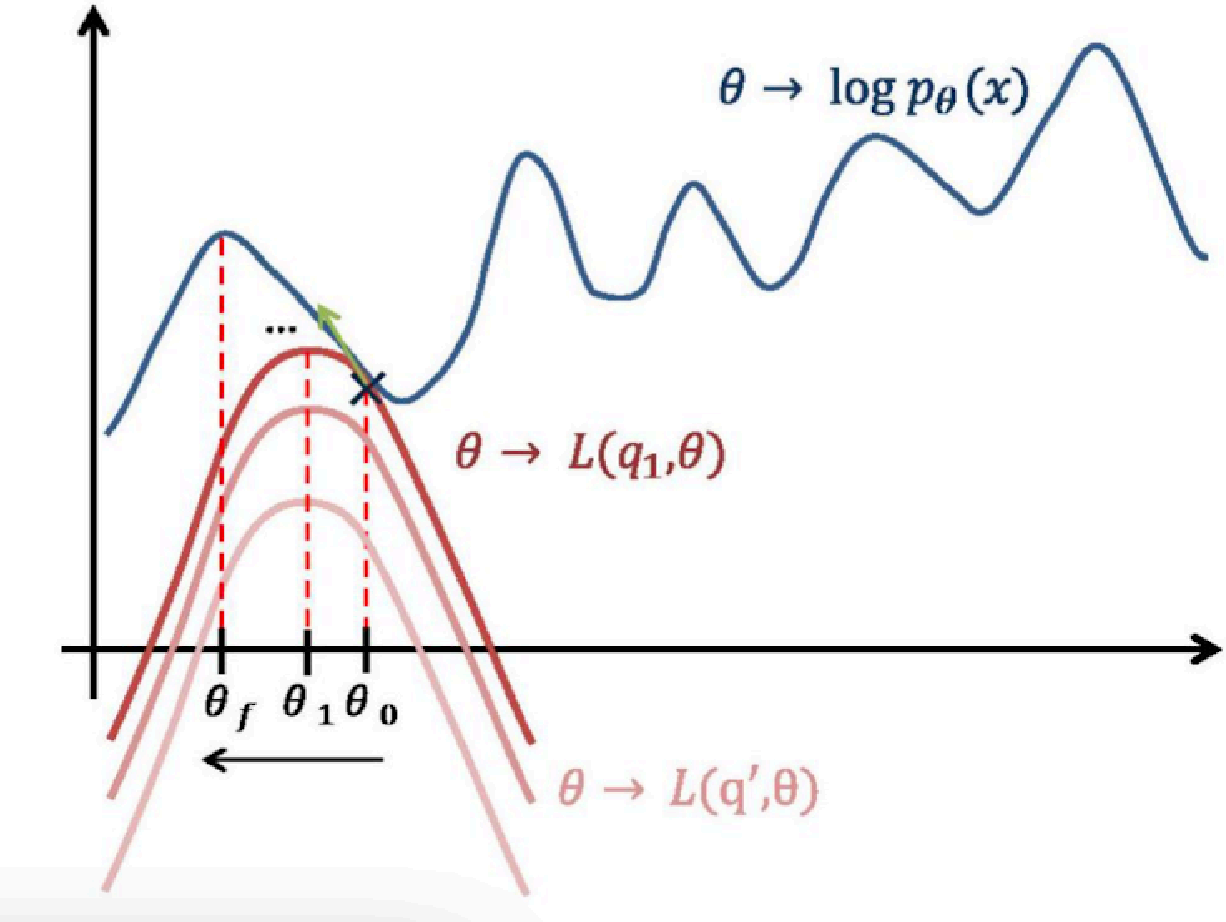
- $a(i)$ = average distance to other points in the same cluster
- $b(i)$ = average distance to points in the nearest different cluster
- Silhouette value $s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))}$

Mark only one oval.

- ( ) Close to 1, indicating strong cluster assignment
- ( ) Close to 0, indicating boundary position
- ( ) Close to -1, indicating misclassification

11. When using the Expectation-Maximization (EM) algorithm for Gaussian Mixture Models, which statement accurately describes its theoretical guarantee?

1 point



Mark only one oval.

- ( ) The EM algorithm converges to the global maximum of the marginal likelihood.
- ( ) The EM algorithm increases the evidence lower bound (denoted as L(q, θ)) at each iteration.
- ( ) It guarantees the optimal number of mixture components

12. In the programming session, after computing the distance matrix, applying PCA, and then using GMMs, we apply Bayes' rule to compute posterior probabilities. What do these represent? *

1 point

Mark only one oval.

- ( ) Hard cluster assignments, like K-means does
- ( ) Soft cluster assignments — probabilities that each point belongs to each cluster
- ( ) A transformation of the distance matrix into centroids

Questions ?

13. Is the pace of the course appropriate? Would a probability recap during office hours be useful?