

## **1. Model Evaluation, Selection and Interpretability (50 points)**

1. (10 points) Compare the bias and variance characteristics of the following cross-validation methods. a) Holdout method with 70% training, 30% testing b) 10-fold cross-validation c) Leave-one-out cross-validation. Which of these will have the highest bias and which will have the highest variance. **Explain your answer in 3-4 sentences.**

2. (15 points) You're using cross-validation to select the best hyperparameters for your model. A colleague suggests using the same cross-validation folds for both hyperparameter tuning and final model evaluation. What's the problem with this approach, and what should you do instead? **Explain your answer 3-5 sentences.**

3. (10 points) A hospital is implementing an ML model to assist doctors in diagnosing *rare* diseases. They have multiple models to choose from, each with different strengths. Why might the precision-recall curve be more informative than the ROC curve in this case? **Explain your answer 3-5 sentences.**

4. (15 points) A ML model for diagnosing a *rare* disease is showing accuracy that surpasses even experienced doctors. While this seems impressive, the medical team is skeptical and wants to understand how the model is achieving this performance.

1. (5 points) Which model approximation technique would you use to get an interpretable version of this complex model? Why?
2. (10 points) Your chosen method shows the model relying heavily on very particular combinations of symptoms. What potential issue in the training data or process might this indicate?

## 2. Ensembles (50 points)

5. (10 points) Excited by the ensembles lecture, you created a *bagging* model with decision tree as base classifier, expecting it to outperform a single decision tree. However, the ensemble's performance is only marginally better than the individual tree. What could be causing the ensemble to underperform? How would you investigate if this is due to high bias or high correlation among the trees? **Explain your answer 3-5 sentences.**

6. (40 points) Please evaluate the following 4 scenarios and describe in **2-3 sentences** (and not more!) whether this is a good or bad idea and **why**.

(10 points) a. Using Boosting when the data has lot of outliers.

(10 points) b. Using Linear SVMs as base classifier for Bagging

(10 points) c. Using Boosting on imbalanced data.

(10 points) d. Using Boosting when the data is noisy. (A noisy dataset refers to data that contains a significant amount of irrelevant information or inaccuracies in feature values.)