1. Did not screenshot the first question
2. **Boosting Classification**

**For the classification problem, Boosting is widely used because**

Pick one option:

1. Diversity can be boosted because all ensemble members are created in parallel
2. Residual error has been minimized, which reduces the margin distribution
3. It attempts to maximize the margins on the training data
4. None of the above

For this one I may have selected A in confusion of my notes on bagging but I’m now sure it’s B.

1. **Random forest**

**How does Random Forest algorithm introduce randomness?**

Pick one or more options:

1. By creating bootstrap samples
2. By selecting a random subset of feature for each tree -
3. By selecting a random subset of feature for each split
4. By using random hyperparameters for each tree

I thought this was a trick question and couldn’t decide between A & C, so I took into account the keyword “introduce” as emphasizing the first step, and chose A. I now see the option to choose both.

1. **Underfit GBoost**

**You have trained a gradient booster model which you find is underfitting. What parameter can you consider decreasing?**

Pick one option:

1. The depth of the tree
2. The number of leaves
3. The proportion of objects used to fit the tree
4. The minimum amount of objects per leaf

I chose D. in this one, interpreting min\_samples\_split to reference leaves where too high of values leads to underfitting. I see the min\_samples\_leaf is separate so this referred to objects. C.?

1. **Complex Function**

**Consider the case of a binary classification problem in which the response is sampled from a non-linear complex function, which of the following algorithms has the potential to perform the best?**

Pick one option:

1. Logistic Regression
2. K-nearest neighbors
3. Linear Discriminant Analysis
4. Quadratic Discriminant Analysis

I selected A which may have been the worst answer. I’m almost sure it’s B as input for the others seems dependent on prior classified output.

1. **Nuclear Missile**

**You are building a nuclear missile detection model which is a binary classification algorithm on radar images. What metric is especially important to consider?**

Pick one option:

1. Recall
2. Specificity
3. F1 Score
4. Accuracy

Initially I wanted to choose C. that takes into account both recall and precision but reasoned with A as a metric that would be of first priority to optimize before resultant F1 score.

1. **Defective Items**

**We want to investigate the proportion of defective items in a production line \*theta\*. Assume a uniform prior for \*theta\*. We get 30 new samples where 3 items are defective. What is the prior mean and the posterior mean?**

Pick one option:

1. Prior mean: 0.5 Posterior mean: .25
2. Prior mean: 0.5 Posterior mean: .125
3. Prior mean: 0.25 Posterior mean: .5
4. Prior mean: 0.125 Posterior mean: .5

I answered this with basic deduced reasoning of prior (before) and posterior (after) and that uniform prior denotes consistency in exponential relevance to posterior means. I chose D. but on second look I’m almost positive the values need to be switched. B.

1. **Flu Season**

**Before the flu season about 20% of people will get the flu vaccine. For those who get the vaccine there is still a 1% chance that they may get the flu. For those who did not get the vaccine there is a higher chance they may get the flu, which is 5%. Now we have a flu patient, what is the probability that he has received a vaccine?**

Pick one option:  
A. 3.5%

B. 4.76%

C. 5.1%

D. 6.3%

Answer is B.

1%20 =.2 , 5%80 =4, 4.2/100 get flu, .2/4.2\*100 = 4.76% chance person with flu is vaccinated

I timed out on the coding question, syntax was confusing. I’ve since come up with an interpretation but did not screenshot constraints so current working model uses initial input values.