3. **(5 points)** A classifier trained on more training data is more likely to over-fit (True / False)

False: The classifier that is trained on more training data will less likely overfit. This is because training accuracy decreases with more training data.

4. **(3 points)** Passive learning is label-efficient when compared to active learning protocol (True/False)

False: The active learning protocol is more efficient because the process involves intelligently selecting the inputs and asking for labels.

5. **(5 points)** If you are given m data points, and use half for training and half for testing, the difference between training error and testing error decreases as m increases. (True/False) Please provide one sentence justification.

True: Intuitively as more data comes in ($m\uparrow$) training error($\uparrow$) increases and test error decreases($\downarrow$). This is mostly because the model will be refined at testing stage (Accurate more)

- **Frequent Itemset and Association Rule Mining (30 points)**

6. **(6 points)** Suppose the support of $\{A\}$ is 5, support of $\{B\}$ is 7, support of $\{A,B\}$ is 4, support of $\{B,C\}$ is 3, support of $\{A,C\}$ is 4, and support of $\{A,B,C\}$ is 2. What is the confidence of following association rules?

6.1 $A \Rightarrow \{B,C\}$

$$\frac{Sup(A,B,C)}{Sup(A)} = \boxed{\frac{2}{5}}$$

6.2 $\{A,B\} \Rightarrow C$

$$\frac{Sup(A,B,C)}{Sup(A,B)} = \frac{2}{4} = \boxed{1/2}$$

Sup(A) = 5
Sup(B) = 7
Sup(A,B) = 4
Sup(B,C) = 3
Sup(A,C) = 4
Sup(A,B,C) = 2

7. **(5 points)** Describe the key property that is exploited by the Apriori algorithm for efficiently computing the frequent itemsets.

Monotonicity property which states that if a set of items appears at least 5 times, so does every subset of the set.

(The negation is also true)