FIRST NAME: .....

LAST NAME: .....

ID NUMBER: .....

1. The goal of supervised learning is:
   (a) to learn useful features
   (b) to learn a model with low generalization error.
   (c) to learn a model with low training error.
   (d) to learn a neural network
   (e) none of the above
2. If you flip 10 times a coin that has probability 0.25 to give tail, the probability that you obtain 10 heads is:
   (a) 0.25×10
   (b) $(1 - 0.25) \times 10$
   (c) $(1 - 0.25)^{10}$
   (d) $(0.25)^{10}$
   (e) none of the above
3. Let $\mathcal{X}, \mathcal{Y}, \mathcal{D}, \ell(\mathbf{x}, y), \mathcal{H}, h, S$ defined as usual during the course. The definition of *training* error $L_S(h)$ is:
   (a) $L_S(h) = \mathbf{E}_{\mathbf{x},y\sim\mathcal{S}}[(h(\mathbf{x}) - y)^2]$
   (b) $L_S(h) = \frac{1}{|S|}\sum_{i=1}^{|S|}(h(\mathbf{x}_i) - y_i)^2$
   (c) $L_S(h) = \mathbf{E}_{\mathbf{x},y\sim\mathcal{S}}[\ell(h, (\mathbf{x}, y))]$
   (d) $L_S(h) = \frac{1}{|S|}\sum_{i=1}^{|S|}\ell(h, (\mathbf{x}_i, y_i))$
   (e) none of the above
4. Let $\mathcal{X}, \mathcal{Y}, \mathcal{D}, \ell(\mathbf{x}, y), \mathcal{H}, h, S$ defined as usual during the course. The definition of *generalization* error $L_{\mathcal{D}}(h)$ is:
   (a) $L_{\mathcal{D}}(h) = \mathbf{E}_{\mathbf{x},y\sim\mathcal{D}}[(h(\mathbf{x}) - y)^2]$
   (b) $L_{\mathcal{D}}(h) = \frac{1}{|\mathcal{D}|}\sum_{i=1}^{|\mathcal{D}|}(h(\mathbf{x}_i) - y_i)^2$
   (c) $L_{\mathcal{D}}(h) = \mathbf{E}_{\mathbf{x},y\sim\mathcal{D}}[\ell(h, (\mathbf{x}, y))]$
   (d) $L_{\mathcal{D}}(h) = \frac{1}{|\mathcal{D}|}\sum_{i=1}^{|\mathcal{D}|}\ell(h, (\mathbf{x}_i, y_i))$
   (e) none of the above
5. To use the ERM approach means that we learn a model by:
   (a) finding the hypothesis with smallest training error
   (b) finding the hypothesis with smallest generalization error
   (c) finding the hypothesis with smallest complexity
   (d) finding the hypothesis that minimizes the expected regularization
   (e) none of the above
6. The difference between classification and regression is given by:
   (a) the type of models you can use (e.g., SVM vs linear models)
   (b) the loss function ℓ you can use
   (c) the approach used to find the model
   (d) the label set $\mathcal{Y}$
   (e) none of the above
7. What does "overfitting" refer to?
   (a) Learning a model that has perfect accuracy on all datasets
   (b) Learning a model that is too simple
   (c) Learning a model that performs well on training data but poorly on new data
   (d) Failing to converge during training
   (e) none of the above

8. The realizability assumption is defined as:
   (a) there exists $h^* \in \mathcal{H}$ with $L_S(h) = 0$
   (b) there exists $h^* \in \mathcal{H}$ with $L_{\mathcal{D}}(h) = 0$
   (c) ERM finds $h^* \in \mathcal{H}$ such that $L_S(h) = 0$
   (d) ERM finds $h^* \in \mathcal{H}$ such that $L_{\mathcal{D}}(h) = 0$
   (e) none of the above
9. For an hypothesis class $\mathcal{H}$, being PAC learnable with respect to a loss function ℓ, means:
   (a) that for all $\mathcal{D}$ we can find the best h in $\mathcal{H}$, with enough data
   (b) that for some $\mathcal{D}$ we can find the best h in $\mathcal{H}$, independently of the amount of data
   (c) that for all $\mathcal{D}$ we can find the best h in $\mathcal{H}$, independently of the amount of data
   (d) that for some $\mathcal{D}$ we can find the best h in $\mathcal{H}$, with enough data
   (e) none of the above
10. What is the main idea behind the concept of bias-complexity trade-off?
    (a) Balancing the trade-off between model simplicity and interpretability
    (b) Balancing the trade-off between accuracy and training time
    (c) Balancing the trade-off between estimation error and approximation error
    (d) Balancing the trade-off between feature selection and feature engineering
    (e) none of the above
11. What is the purpose of regularization?
    (a) To reduce model complexity so to prevent overfitting
    (b) To increase model complexity so to prevent overfitting
    (c) To improve training speed
    (d) To eliminate bias in the model
    (e) none of the above
12. What is the purpose of validation?
    (a) To train the model on multiple datasets
    (b) To assess the model's performance on the training set
    (c) To compare different models on unseen data
    (d) To obtain a good estimate of the generalization error
    (e) none of the above
13. What is a main difference between SVMs and linear models?
    (a) there is no difference
    (b) SVM can be used to learn models that are polynomial in the features
    (c) SVMs can be used only for linearly separable data
    (d) SVMs consider the margin of the model, linear models do not
    (e) none of the above
14. The VC dimension is a measure of:
    (a) the dimension of each point in a dataset
    (b) the complexity of an hypothesis class
    (c) the number of features in a model
    (d) the generalizability of an hypothesis
    (e) none of the above
15. What is the primary goal of clustering?
    (a) predicting a continuous target variable
    (b) assigning labels to data points
    (c) identifying patterns in unlabeled data
    (d) classifying data into predefined categories
    (e) none of the above