Final: Multiple answer

First Name, SURNAME:

[50 pts] Multiple Answer

Fill in the bubbles for **ALL correct choices**: there may be more than one correct choice, but there is always at least one correct choice. **NO partial credit**: the set of all correct answers must be checked.

1. (5 points) Recall that in subset selection, we attempt to identify poorly predictive features and ignore them. Which of the following are reasons why we may seek to drop features available to our model?
   - ◯ A: To reduce model bias
   - ◯ B: To reduce model variance
   - ◯ C: To increase speed of prediction on test points
   - ◯ D: To improve model interpretability

2. (5 points) Let $\mathbf{X}$ be a real-valued $n \times p$ matrix. Let $\mathbf{\Lambda}$ be a diagonal, real-valued $n \times n$ matrix whose diagonal entries are all positive. Which of the following are true of the matrix product $\mathbf{M} = \mathbf{X}^\top \mathbf{\Lambda} \mathbf{X}$?
   - ◯ A: $\mathbf{M}$ could have negative eigenvalues
   - ◯ B: $\mathbf{M}$ could have eigenvalues equal to zero
   - ◯ C: $\mathbf{M}$ could have positive eigenvalues
   - ◯ D: the eigenvalues of $\mathbf{M}$ are the values on the diagonal of $\mathbf{\Lambda}$

3. (5 points) Select the correct statements about AdaBoost.
   - ◯ A: "Ada" stands for "adaptive", as the metalearner adapts to the performance of its learners
   - ◯ B: AdaBoost works best with support vector machines
   - ◯ C: At test/classification time, AdaBoost computes a weighted sum of predictions
   - ◯ D: AdaBoost can transform any set of classifiers to a classifier with better training accuracy

4. (5 points) Consider the optimization problem of finding $\mathbf{w} \in \mathbb{R}^r$ that minimizes $\|Y - XP\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_2^2$, where $\lambda > 0$, $X \in \mathbb{R}^{n \times d}$ a design matrix, $Y \in \mathbb{R}^n$ is a vector of labels, and $P \in \mathbb{R}^{d \times r}$ is an arbitrary matrix with $r < d$. This problem has one unique solution. Which of the following is that one unique solution?

○ $\mathbf{w}^* = (P^\top X^\top X P + \lambda I)^{-1} P^\top X^\top Y$ ○ $\mathbf{w}^* = (P^\top X^\top X P + \lambda I)^{-1} X^\top Y$

○ $\mathbf{w}^* = (X^\top P^\top P X + \lambda I)^{-1} P^\top X^\top Y$ ○ $\mathbf{w}^* = (X^\top P^\top P X + \lambda I)^{-1} X^\top Y$

---

5. (5 points) Which of the following are benefits of using convolutional neural networks—as opposed to fully connected ones— for image recognition tasks?

○ A: The ability to express a wider variety of more complicated functions of the input features

○ B: Fewer model architecture hyperparameters for the designer to select

○ C: Enables the network to more easily learn and recognize features regardless of their position in the image

○ D: Typically requires less data to train well

---

6. (5 points) Select the correct statements about principal component analysis (PCA).

○ A: PCA is a method of dimensionality reduction

○ B: If we select only the first direction (a one-dimensional subspace) to represent the data, the sample variance of the projected points is zero if and only if the original sample points are all identical

○ C: The orthogonal projection of a point $\mathbf{x}$ onto a unit direction vector $\mathbf{w}$ is $(\mathbf{x}^\top \mathbf{w})\mathbf{w}$

○ D: If we select only the first direction (a one-dimensional subspace) to represent the data, PCA chooses the eigenvector of the sample covariance matrix that corresponds to the least eigenvalue

---

7. (5 points) Suppose we have a feature map Φ and a kernel matrix $\mathbf{K} = (k(X_i, X_j))_{i,j=1}^n$ with $k(X_i, X_j) = \Phi(X_i)^\top \Phi(X_j)$. Select the true statements about kernels.

○ If there are n sample points of dimension d, it takes $O(nd)$ time to compute the kernel matrix

○ The kernel trick implies we do not compute $\Phi(X_i)$ explicitly for any sample point X_i

○ $\|\Phi(X_i) - \Phi(X_j)\|_2^2$ can be computed using exclusively inner products

○ Running times of kernel algorithms do not depend on the dimension D of the feature space $\Phi(.)$

---

8. (5 points) Suppose we have a covariance matrix, $\begin{pmatrix} 5 & a \\ a & 4 \end{pmatrix}$. The set of values that the scalar a can take on such that the matrix is a valid covariance matrix is:

- ○ $a \in \mathbb{R}$
- ○ $a \geq 0$
- ○ $-9 < a < 9$
- ○ None of the above
- ○ $-9 \leq a \leq 9$
- ○ $-\sqrt{20} < a < \sqrt{20}$
- ○ $-\sqrt{20} \leq a \leq \sqrt{20}$

---

9. (5 points) Consider least-squares linear regression with a design matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ and target $Y \in \mathbb{R}^n$. If the solution to the least-squares problem is unique, which of the following must be true?

- ○ A: $\text{rank}(\mathbf{X}) = p$
- ○ B: $\text{rank}(\mathbf{X}^\top) = p$
- ○ C: $n \leq p$
- ○ D: $p \leq n$

---

10. (5 points) Suppose your training set for two-class classification in **one dimension** ($d = 1$; $x_i \in \mathbb{R}$) contains three sample points: point $x_1 = 3$ with label $y_1 = 1$, point $x_2 = 1$ with label $y_2 = 1$, and point $x_3 = -1$ with label $y_3 = -1$. What are the values of w and b given by a **hard-margin SVM**?

- ○ A: $w = 1, b = 1$
- ○ B: $w = 0, b = 1$
- ○ C: $w = 1, b = 0$
- ○ D: $w = \infty, b = 0$

Left Blanck

Final

- The exam is open book, open notes **for material on paper.**
- **Electronic devices are forbidden on your person,** including cell phones, tablets, headphones, and laptops. Leave your cell phone off and in a bag; it should not be visible during the exam.
- Fill in the "Multiple Answers" sheet directly and return it with your copy. Please enter your name in the appropriate field.
- You will submit your answers to the written questions by writing them on paper by hand.
- Please write your name at the top of each page of your written answers. (You may do this before the exam.) Clearly label all written questions and all subparts of each written question.
- You have 2 hours (unless you are in the DSP program and have a larger time allowance).
- The total number of points is 100. There are 10 multiple choice questions worth 5 points each, and 4 written questions worth a total of 50 points.