"Machine Learning and Computational Statistics"

9th Homework

(A) Support vector machines (SVM)

Exercise 1 (mult. choice question): Consider a two-class classification task, where the involved classes are labeled as $+1$ and -1. Assume that a data set $X = \{(y_n, \boldsymbol{x}_n),\ \boldsymbol{x}_n \in R^l,\ y_n \in \{-1, +1\},\ n = 1, \ldots, N\}$ is available for this task. Which of the following statements is/are true?

1. The two classes are linearly separable if there exists a hyperplane that leaves all points of X from class +1 (resp. -1) on its positive (resp. negative) side.

2. The two classes are linearly separable if the probability density functions (pdfs) that model the two classes have disjoint support sets.

3. The two classes are not linearly separable if there exists only a single line that leaves all points of X from class +1 (resp. -1) on its positive (resp. negative) side.

4. If the pdfs that model the two classes have not disjoint supports, then the data points from class $+1$ in X are definitely not linearly separable from the data points from class -1 in X.

Hint: The support set of a pdf function $p(\boldsymbol{x})$ from R^l to R is defined as $\text{supp}(p) = \{\boldsymbol{x} \in R^l : p(\boldsymbol{x}) \neq 0\}$.

Exercise 2 (mult. choice question): Consider a two-class classification task, where the involved classes are labeled as $+1$ and -1. Assume that a data set $X = \{(y_n, \boldsymbol{x}_n),\ \boldsymbol{x}_n \in R^l,\ y_n \in \{-1, +1\},\ n = 1, \ldots, N\}$ is available for this task. Which of the following statements is/are true?

1. If the support sets of the pdfs that model the two classes are not disjoint, the classification problem is definitely not-linearly separable.

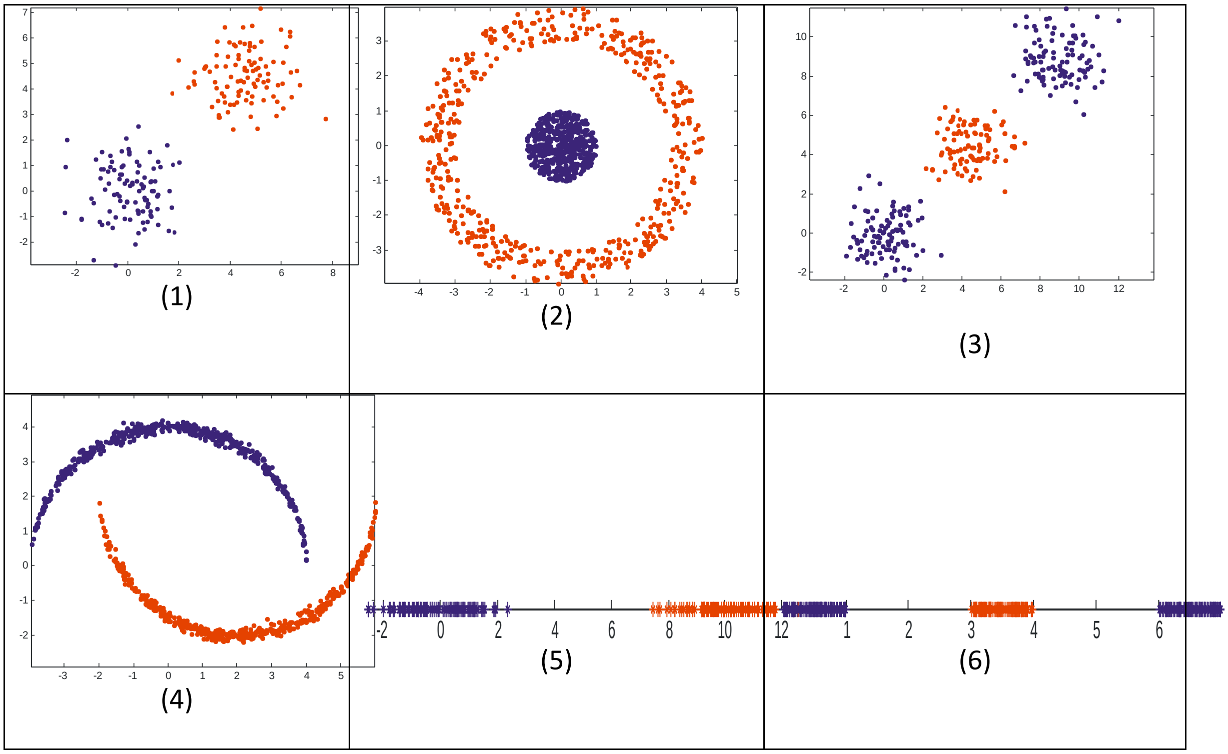
2. If there exists a hyperplane that leaves all points from class +1 (resp. -1) on its positive (resp. negative) side, then this is unique.

3. If a hyperplane passes through the region where the points of class +1 lie (that is, it leaves some of them on its negative side), this implies that the classification problem is not linearly separable.

4. If a hyperplane H classifies correctly all the points of X, there exists a hyperplane H' parallel to H that does the same.

Hint: The support set of a pdf function $p(\boldsymbol{x})$ from R^l to R is defined as $\text{supp}(p) = \{\boldsymbol{x} \in R^l : p(\boldsymbol{x}) \neq 0\}$.

Exercise 3 (mult. choice question): The following figures depict the data points associated with one and two-dimensional two-class classification tasks. In which of these cases the classes are linearly separable?



Exercise 4 (mult. choice question): Consider the equation $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$, where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_l]^T \in R^l$ and $\boldsymbol{x} = [x_1, \dots, x_l]^T \in R^l$ and $\theta_0 \in R$. Which of the following statements is/are true?

1. For a specific $\boldsymbol{\theta}$**,** the equation $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$ models an infinite set of parallel hyperplanes in the R^l space, as θ_0 varies.

2. For $\theta_0 = 0$, the equation $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$ models all the hyperplanes in the R^l space that do not pass through the origin.

3. If $\theta_0 \neq 0$, at least one of the family of hyperplanes, defined by the equation $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$, passes through the origin.

4. For fixed $\boldsymbol{\theta}$ and θ_0, if for a specific point $\boldsymbol{x} \in R^l$ it is $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 > 0$, then $\boldsymbol{x}$ lies on the positive side of the hyperplane H defined by the equation $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$.

5. For fixed $\boldsymbol{\theta}$, all the hyperplanes that are modeled by $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$, are parallel to the hyperplane defined by the equation $\boldsymbol{\theta}^T\boldsymbol{x} = 0$

Exercise 5 (mult. choice question): Consider two hyperplanes H and H' in R^l, defined by the equations $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$ and $\boldsymbol{\theta}'^T\boldsymbol{x} + \theta'_0 = 0$, respectively, where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_l]^T$, $\boldsymbol{\theta}' = [\theta'_1, \dots, \theta'_l]^T$ and $\boldsymbol{x} = [x_1, \dots, x_l]^T$. Which of the following statements is/are true?

1. H is parallel to the vector $\boldsymbol{\theta}$.

2. If $\boldsymbol{\theta} = \boldsymbol{\theta}'$, then H' is parallel to H.

3. If $\theta_0 \neq \theta'_0$, then H' is definitely not parallel to H. .

4. If the intersection of H and H' is not empty, then $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ are not collinear (i.e., $\boldsymbol{\theta} \neq a\boldsymbol{\theta}', a \in R$)

Exercise 6 (mult. choice question): The (Euclidean) distance of a point $\boldsymbol{x} \in R^l$ from the hyperplane H: $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$, where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_l]^T$ and $\boldsymbol{x} = [x_1, \dots, x_l]^T$, is $z = \frac{|\boldsymbol{\theta}^T\boldsymbol{x}+\theta_0|}{||\boldsymbol{\theta}||}$, where $||\boldsymbol{\theta}|| = \sqrt{{\theta_1}^2 + \cdots {\theta_l}^2}$. Which of the following statements is/are true?

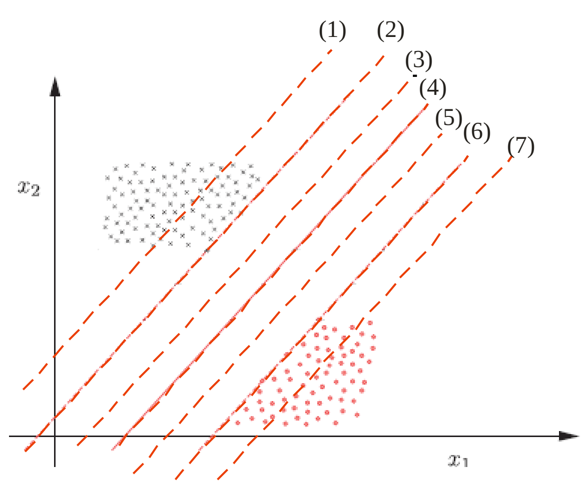
1. If for a specific point $\boldsymbol{x}$ it is $z = 0$, then $\boldsymbol{x} \in H$.

2. If the distances z_1, z_2 of the points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in R^l$ from H are equal, then the points lie necessarily on a line that is parallel to H.

3. If a point $\boldsymbol{x}$ lies on the negative side of H, its distance z from it is negative.

4. Two points that are symmetric with respect to H, have equal distances from it.

Exercise 7 (mult. choice question): Consider a two-class two-dimensional classification task, for which a data set X is available, whose points are shown in the figure to the right (red-colored points belong to class $+1$ and black-colored points belong to class -1). The lines shown in the figure are parallel to each other. Which of the following statements is/are true?



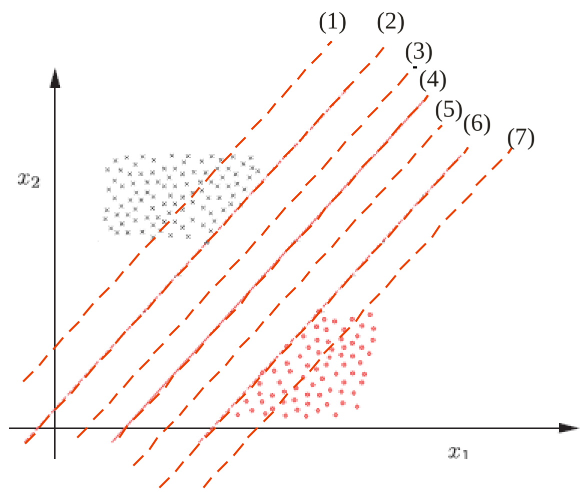
1. The classes are linearly separable.

2. All the lines correspond to the same θ_0.

3. Not all the lines have the same direction.

4. Line (5) separates the class $+1$ data points in X from the class -1 data points in X.

Exercise 8 (mult. choice question): Consider a two-class two-dimensional classification task, for which a data set X is available, whose points are shown in the figure to the right (red-colored points belong to class $+1$ and black-colored points belong to class -1). The lines shown in the figure are parallel to each other. Which of the following statements is/are true?



1. Line (1) separates perfectly the two classes.

2. The distances of the nearest to line (3) points from the two classes are equal.

3. The product of the distances of the two nearest to line (4) points from the two classes is larger than that associated with Line (3).

4. The product of the distances of the two nearest to line (2) points from the two classes is zero.

<u>*Hint:*</u> Among all pairs of positive numbers (a, b) with constant sum, the pair that maximizes the product $a \cdot b$ is the (a, a).

Exercise 9: Consider a two-dimensional, two-class classification task, for which a data set X is available. The involved classes are linearly separable. Recall that each vector (direction) $\boldsymbol{\theta}$ is associated with an infinite set, $S_{\boldsymbol{\theta}}$, of parallel to each other hyperplanes, that are perpendicular to $\boldsymbol{\theta}$. Which of the following statements is/are true?

1. There are directions $\boldsymbol{\theta}$ so that none of the hyperplanes in $S_{\boldsymbol{\theta}}$ manages to separate the classes.

2. If, for a specific direction $\boldsymbol{\theta}$, there exists at least one hyperplane in $S_{\boldsymbol{\theta}}$ that separates the two classes, then all the hypeprlanes in $S_{\boldsymbol{\theta}}$ separate the two classes.

3. If, for a specific direction $\boldsymbol{\theta}$, there exists at least one hyperplane in $S_{\boldsymbol{\theta}}$ that separates the two classes, then there exists an infinite set $S'_{\boldsymbol{\theta}} \subset S_{\boldsymbol{\theta}}$ of hyperplanes that separate the two classes.

4. For a specific direction $\boldsymbol{\theta}$, there exists exactly one hyperplane H in $S_{\boldsymbol{\theta}}$ that is equidistant from the two nearest to it points from the two classes.

5. For a specific direction $\boldsymbol{\theta}$, there exists no hyperplane H in $S_{\boldsymbol{\theta}}$ that solves the problem, whose distance from the nearest point of class $+1$ is zero.

Exercise 10 (mult. choice question): Consider a two-class two-dimensional classification task, for which a data set X is available. The involved classes are linearly separable. Consider two non-collinear vectors (directions) $\boldsymbol{\theta}_i$, $i = 1,2$, where each one is associated with an infinite set, $S_{\boldsymbol{\theta}_i}$, of parallel to each other hyperplanes, that are perpendicular to $\boldsymbol{\theta}_i$. Assume that there are hyperplanes in both $S_{\boldsymbol{\theta}_i}$'s that solve the classification task. Let H_i^1 and H_i^{-1} be the two hyperplanes in $S_{\boldsymbol{\theta}_i}$ that solve the problem and H_i^1 (resp. H_i^{-1}) have zero distance from its closest from class 1 (resp. -1) data point. Which of the following statements is/are true?

1. In certain cases, the sets $S_{\boldsymbol{\theta}_1}$ and $S_{\boldsymbol{\theta}_2}$ coincide.

2. The hyperplanes H_1^1 and H_2^1 do not coincide.

3. The distance d_1 between H_1^1 and H_1^{-1} is, in general, different from the distance d_2 between H_2^1 and H_2^{-1}.

4. The hyperplane that is equidistant from H_1^1 and H_1^{-1} is parallel to the hyperplane that is equidistant from H_2^1 and H_2^{-1}.

Exercise 11 (mult. choice question): Consider the line (ε): $2x_1 + 4x_2 + 3 = 0$ and the lines (ε_1): $2x_1 + 4x_2 + 3 = 2$ and (ε_2): $2x_1 + 4x_2 + 3 = -2$. Which of the following statements is/are true?

1. The lines (ε_1) and (ε_2) are both parallel to (ε).

2. The lines (ε_1) and (ε_2) have a unique intersection point.

3. The distance between (ε) and (ε_1) is less than the distance between (ε) and (ε_2).

4. The lines (ε_1) and (ε_2) can be expressed equivalently by the equations $x_1 + 2x_2 + \frac{3}{2} = 1$ and $x_1 + 2x_2 + \frac{3}{2} = -1$, respectively.

Exercise 12 (mult. choice question): Consider the lines (ε): $2x_1 + 4x_2 + 3 = 0$, (ε_1): $2x_1 + 4x_2 + 3 = 2$, (ε'_1): $x_1 + 2x_2 + \frac{3}{2} = 1$, (ε_2): $2x_1 + 4x_2 + 3 = -2$ and (ε'_2): $x_1 + 2x_2 + \frac{3}{2} = -1$. Let A be a point with coordinates $(\tilde{x}_1, \tilde{x}_2)$ on (ε) and let d, d_1, d'_1, d_2, d'_2 be the distances of A from (ε), (ε_1), (ε'_1), (ε_2) and (ε'_2), respectively. Which of the following statements is/are true?

1. $d = 0, d_1 = \frac{\sqrt{5}}{5}, d'_1 = \frac{\sqrt{5}}{5}, d_2 = \frac{\sqrt{5}}{5}, d'_2 = \frac{\sqrt{5}}{5}$.

2. $d = 0, d_1 = \frac{\sqrt{5}}{5}, d'_1 = \frac{1}{\sqrt{5}}, d_2 = \frac{\sqrt{5}}{5}, d'_2 = \frac{1}{\sqrt{5}}$.

3. $d = 0, d_1 = \frac{\sqrt{5}}{5}, d'_1 = \frac{\sqrt{5}}{5}, d_2 = -\frac{\sqrt{5}}{5}, d'_2 = -\frac{\sqrt{5}}{5}$.

4. $d = 0, d_1 = \frac{\sqrt{5}}{5}, d'_1 = \frac{\sqrt{5}}{10}, d_2 = \frac{\sqrt{5}}{5}, d'_2 = \frac{\sqrt{5}}{10}$.

Hint: The distance of a point $\boldsymbol{x} \in R^l$ from the hyperplane H: $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$, where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_l]^T$ and $\boldsymbol{x} = [x_1, \dots, x_l]^T$, is $z = \frac{|\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0|}{||\boldsymbol{\theta}||}$, where $||\boldsymbol{\theta}|| = \sqrt{\theta_1{}^2 + \cdots \theta_l{}^2}$.

Exercise 13 (mult. choice question): Consider the hyperplanes (H): $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$, (H_1): $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = c$, (H'_1): $\left(\frac{1}{c}\boldsymbol{\theta}\right)^T \boldsymbol{x} + \frac{\theta_0}{c} = 1$, (H_2): $\boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = -c$ and (H'_2): $\left(\frac{1}{c}\boldsymbol{\theta}\right)^T \boldsymbol{x} + \frac{\theta_0}{c} = -1$. Let A be a point on (H) with coordinates $(\tilde{x}_1, \tilde{x}_2)$, and let d, d_1, d'_1, d_2, d'_2 be the distances of A from (H), (H_1), (H'_1), (H_2) and (H'_2), respectively. Which of the following statements is/are true?

1. $d > 0$.

2. $d_1 = d'_1$ and $d_2 = d'_2$.

3. $d_1 < d_2$.

4. $d'_1 = d'_2$.

5. The distance between (H'_1), (H_2) equals to $2 \cdot d_1$.

Hint: The distance of a point $\boldsymbol{x} \in R^l$ from the hyperplane $H: \boldsymbol{\theta}^T \boldsymbol{x} + \theta_0 = 0$, where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_l]^T$ and $\boldsymbol{x} = [x_1, \dots, x_l]^T$, is $z = \frac{|\boldsymbol{\theta}^T \boldsymbol{x} + \theta_0|}{||\boldsymbol{\theta}||}$, where $||\boldsymbol{\theta}|| = \sqrt{\theta_1{}^2 + \cdots \theta_l{}^2}$.

Exercise 14 (mult. choice question): Consider a two-class classification task where the two-classes, labeled as $+1$ and -1, are linearly separable. As a consequence, there exists an infinite number of hyperplanes $\boldsymbol{\theta}^T \boldsymbol{x} + \theta_0 = 0$ of various directions $\boldsymbol{\theta}$ that separate the two classes. For such a direction we consider the pair of (parallel) hyperplanes $H_{\boldsymbol{\theta}}{}^+$ and $H_{\boldsymbol{\theta}}{}^-$ such that $H_{\boldsymbol{\theta}}{}^+$ (resp. $H_{\boldsymbol{\theta}}{}^-$) leaves all the points from class +1 (resp. -1) on its positive (resp. negative) side and passes through at least one of them. The distance of $H_{\boldsymbol{\theta}}{}^+$ and $H_{\boldsymbol{\theta}}{}^-$ defines the margin associated with the direction $\boldsymbol{\theta}$. Which of the following statements is/are true?

1. Some directions $\boldsymbol{\theta}$ may be associated with more than one margins.

2. If $H_{\boldsymbol{\theta}}{}^+: \boldsymbol{\theta}^T \boldsymbol{x} + \theta_0 = c$ and $H_{\boldsymbol{\theta}}{}^-: \boldsymbol{\theta}^T \boldsymbol{x} + \theta_0 = -c$, the margin associated with $\boldsymbol{\theta}$ is $\frac{2 \cdot c}{||\boldsymbol{\theta}||}$.

3. If $H_{\boldsymbol{\theta}}{}^+: \left(\frac{1}{c}\boldsymbol{\theta}\right)^T \boldsymbol{x} + \frac{\theta_0}{c} = 1$ and $H_{\boldsymbol{\theta}}{}^-: \left(\frac{1}{c}\boldsymbol{\theta}\right)^T \boldsymbol{x} + \frac{\theta_0}{c} = -1$, the margin associated with $\boldsymbol{\theta}$ is $\frac{2}{\left\|\frac{1}{c}\boldsymbol{\theta}\right\|}$.

4. The margin associated with a direction $\boldsymbol{\theta}$ remains positive even if $H_{\boldsymbol{\theta}}{}^+$ and $H_{\boldsymbol{\theta}}{}^-$ coincide.

Exercise 15 (mult. choice question): Consider a two-class classification task where the (data of the) two classes are linearly separable. In the framework of support vector machines (SVMs) the aim is first to determine the direction $\boldsymbol{\theta}$ whose associated margin is maximum and then to establish among all hyperplanes of direction $\boldsymbol{\theta}$ the one, say (H), that is equidistant from the closest data points from the two classes. Which of the following statements concerning the rationale of SVMs is/are true?

1. It is expected that such a classifier will have better generalization abilities.

2. Data that were not used during the training phase will definitely lie outside the margin associated with $\boldsymbol{\theta}$.

3. The narrower the margin, the more likely is to that data not used during training to be misclassified.

4. It is not guaranteed that a direction with strictly positive margin exists for the above problem.

(B) Generalized linear models

Exercise 16 (mult. choice): Consider a regression task, $\mathrm{y} = g(\mathbf{x}) + \mathrm{e}$, where $\mathbf{x} \in R^l$ is an l-dimensional random vector, y and e are random variables (dependent variable and noise, respectively). Which of the following expressions is/are **not** valid generalized linear estimators $\hat{\mathrm{y}}$ of y (φ_k are preselected (nonlinear) functions)?

1. $\hat{\mathrm{y}} = \theta_0 + \sum_{k=1}^{K} \theta_k \cdot \varphi_k(\mathbf{x})$

2. $\hat{\mathrm{y}} = \theta + \sum_{k=1}^{K} \theta \cdot \varphi_k(\mathbf{x})$

3. $\hat{\mathrm{y}} = \theta_0 + \sum_{k=1}^{K} \theta_k \cdot \varphi_k(\mathbf{x}) \cdot \varphi_1(\mathbf{x})$

4. $\hat{\mathrm{y}} = {\theta_0}^2 + \sum_{k=1}^{K} {\theta_k}^2 \cdot \varphi_k(\mathbf{x})$

5. $\hat{\mathrm{y}} = \sum_{k=1}^{K} \theta_k \cdot \theta_2 \cdot \varphi_k(\mathbf{x})$

Exercise 17 (mult. choice): Consider a regression task, $\mathrm{y} = g(\mathbf{x}) + \mathrm{e}$, where $\mathbf{x} = [\mathrm{x}_1,\ \mathrm{x}_2]^{\boldsymbol{T}} \in R^2$ is a two-dimensional random vector, y and e are random variables (dependent variable and noise, respectively). Which of the following expressions is/are valid second-degree polynomial generalized linear estimators $\hat{\mathrm{y}}$ of y (that is, the preselected functions are monomials up to the second degree and at least one is of second degree)?

1. $\hat{\mathrm{y}} = \theta_0 + \theta_1 \cdot \mathrm{x}_1 + \theta_2 \cdot \mathrm{x}_2 + \theta_3 \cdot \mathrm{x}_1 \cdot \mathrm{x}_2 + \theta_4 \cdot {\mathrm{x}_1}^2 + \theta_5 \cdot {\mathrm{x}_2}^2$

2. $\hat{\mathrm{y}} = \theta_0 + \theta_1 \cdot \mathrm{x}_1 + \theta_2 \cdot \mathrm{x}_2$

3. $\hat{\mathrm{y}} = \theta_0 + \theta_1 \cdot \mathrm{x}_1 + \theta_2 \cdot \mathrm{x}_2 + \theta_3 \cdot \mathrm{x}_1 \cdot \mathrm{x}_2 + \theta_4 \cdot {\mathrm{x}_1}^2 + \theta_5 \cdot \mathrm{x}_1 \cdot {\mathrm{x}_2}^2$

4. $\hat{\mathrm{y}} = \theta_0 + \theta_1 \cdot \mathrm{x}_1 + \theta_2 \cdot \mathrm{x}_2 + \theta_5 \cdot {\mathrm{x}_2}^2$

Exercise 18 (mult. choice): Consider a regression task, $\mathrm{y} = g(\mathrm{x}) + \mathrm{e}$, where $\mathrm{x} \in R$ is a two-dimensional random vector, x, y and e are random variables (independent variable, dependent variable and noise, respectively). Consider the family of a third degree polynomial generalized linear estimators $\hat{\mathrm{y}}$ of y, $\hat{\mathrm{y}} = \theta_0 + \theta_1 \cdot \mathrm{x} + \theta_2 \cdot \mathrm{x}^2 + \theta_3 \cdot \mathrm{x}^3$. If the last expression is to be written in the general form $\hat{\mathrm{y}} = \theta_0 + \sum_{k=1}^{K} \theta_k \cdot \varphi_k(\mathrm{x})$, which of the following is the correct choice for $\varphi_k(\mathrm{x})$'s?

1. $\varphi_1(\mathrm{x}) = 1,\ \varphi_2(\mathrm{x}) = \mathrm{x}^2, \varphi_3(\mathrm{x}) = \mathrm{x}^3$

2. $\varphi_1(\mathrm{x}) = \mathrm{x},\ \varphi_2(\mathrm{x}) = \mathrm{x}^3, \varphi_3(\mathrm{x}) = \mathrm{x}^3$

3. $\varphi_1(\mathrm{x}) = \mathrm{x},\ \varphi_2(\mathrm{x}) = \mathrm{x}^2, \varphi_3(\mathrm{x}) = \mathrm{x}$

4. $\varphi_1(\mathrm{x}) = \mathrm{x},\ \varphi_2(\mathrm{x}) = \mathrm{x}^2, \varphi_3(\mathrm{x}) = \mathrm{x}^3$

Exercise 19 (mult. choice): Consider a regression task, $\mathrm{y} = g(\mathbf{x}) + \mathrm{e}$, where $\mathbf{x} = [\mathrm{x}_1, \dots, \mathrm{x}_l]^{\boldsymbol{T}} \in R^l$ is an l-dimensional random vector, y and e are random variables (dependent variable and noise, respectively). Consider the r-th degree polynomial generalized linear estimators $\hat{\mathrm{y}}$ of y, $\hat{\mathrm{y}} = \theta_0 + \sum_{k=1}^{K} \theta_k \cdot \varphi_k(\mathbf{x})$ (that is, the preselected functions are monomials up to the r-th degree and at least one is of r-th degree), where the number of the involved terms (including θ_0) is $K = \frac{(r+l)!}{r! \cdot l!}$ ($n!$ denotes the factorial of n). Assume that $l = 2$ and $r = 3$ and consider the following two scenarios (a) the general case where the parameters θ_k associated with the φ_k's are allowed to be different from each other and (b) all the terms (monomials) of the same degree are constrained to be multiplied by the same parameter. How many distinct parameters θ_k (at the most) are involved in each scenario?

1. (a) 10, (b) 4

2. (a) 20, (b) 4

3. (a) 4, (b) 10

4. (a) 10, (b) 10

Exercise 20 (mult. choice): Consider the nonlinear regression task of the form $\mathrm{y} = g(\mathbf{x}) + \mathrm{e}$, where $\mathbf{x} = [\mathrm{x}_1, \dots, \mathrm{x}_l]^T$ is an l-dimensional random vector in a compact (closed and bounded) subset of R^l, y and e are random variables (dependent variable and noise, respectively). Consider, also, the generalized linear model of the form $\hat{\mathrm{y}} = \theta_0 + \sum_{k=1}^{K} \theta_k \cdot \varphi_k(\mathbf{x})$, where φ_k are preselected functions. Which of the following is/are true?

1. The basis functions, φ_k, in a generalized linear model are selected independently from the data.

2. The regression problem in the space defined by the mapping $\boldsymbol{x} \to \boldsymbol{\varphi}(\boldsymbol{x})$, where $\boldsymbol{\varphi}(\boldsymbol{x}) = [\varphi_0(\boldsymbol{x}), \dots, \varphi_K(\boldsymbol{x})]^T$, is nonlinear in the parameters.

3. For $l = 1$, the approximation error cannot be smaller than $\frac{1}{K^2}$.

4. The approximation error for the generalized linear models can become arbitrarily small.

5. For a constant approximation error, the number of parameters involved in a polynomial generalized linear estimator of y decreases with the dimensionality l.

Exercise 21 (mult. choice): Consider a continuous real-valued function $y = g(\boldsymbol{x})$, with $\boldsymbol{x}$ lying in a compact subset of R^l. What is the minimum number of monomial terms in a polynomial $p(\boldsymbol{x})$, in order to approximate $g(\boldsymbol{x})$ with approximation error of order 0.01, for (a) $l = 1$ and (b) $l = 2$?

1. (a) 10, (b) 100

2. (a) 100, (b) 10

3. (a) 10, (b) 10

4. (a) 100, (b) 100

Hint: The approximation error cannot be smaller than order of $\left(\frac{1}{K}\right)^{\frac{2}{l}}$.

Exercise 22 (mult. choice): Consider a nonlinear regression task of the form $\mathrm{y} = g(\mathbf{x}) + \mathrm{e}$, where $\mathbf{x} = [\mathrm{x}_1, \dots, \mathrm{x}_l]^T$ is an l-dimensional random vector in a compact subset of R^l, y and e are random variables (dependent variable and noise, respectively). Let, also, $\hat{\mathrm{y}} = \theta_0 + \sum_{k=1}^{K} \theta_k \cdot \varphi_k(\mathbf{x})$ be a generalized linear estimator of y, where φ_k are nonlinear functions. Which of the following statements is/are true?

1. If φ_k are preselected, the lower bound of approximating y by $\hat{\mathrm{y}}$ depends on l.

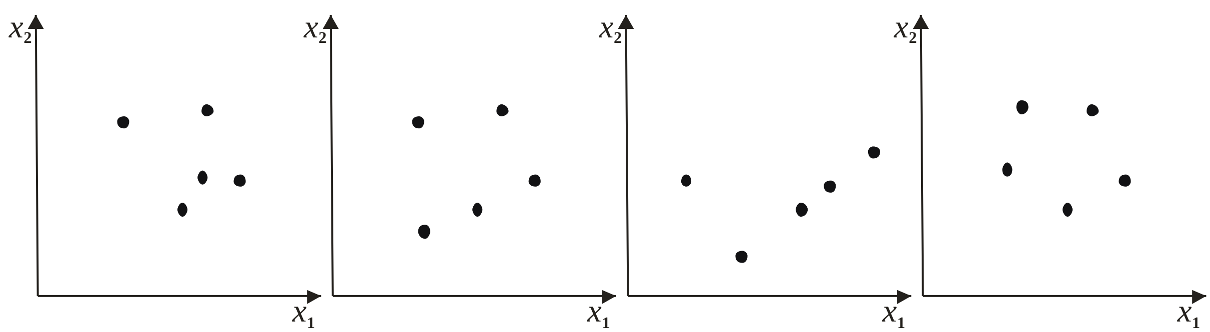
2. If φ_k are data dependent, the lower bound of approximating y by $\hat{\mathrm{y}}$ may not be dependent on l.

3. If φ_k are data dependent, they can only be determined in terms of the joint probability density function of y and $\mathbf{x}$, $p(y, \boldsymbol{x})$.

4. A typical example where φ_k are data dependent are the neural networks.

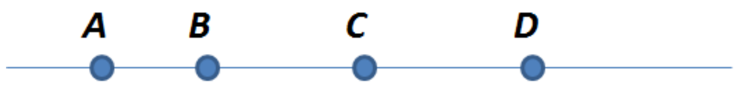
(C) Covers theorem

Exercise 23 (mult. choice): In which of the following cases the data points (in a two-dimensional space) lie in a general position?



(1) (2) (3) (4)

Exercise 24 (mult. choice): Consider the set of $N = 4$ points in the one-dimensional space (line), depicted below. In this case, the dimension of the data space is $l = 1$, so an $(l-1)$-dimensional hyperplane is just a point on the line (data space). There are $2^4 = 16$ ways to group the points into two disjoint groups. Which of the following groupings can be implemented by a point (linear dichotomy) in the line (in other words, given a two-group grouping of the data points, is there a point that leaves on its one side the points of the first group and on its other side the points of the second group)?



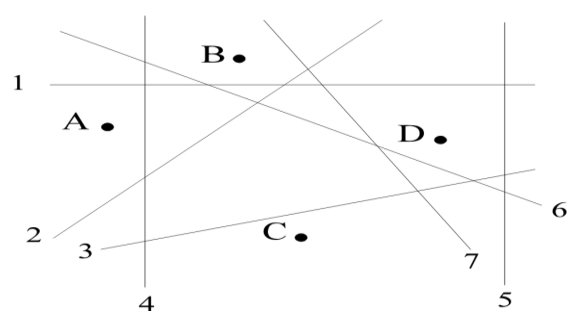
1. $\{A, B\}, \{C, D\}$

2. $\{A\}, \{B, C, D\}$

3. $\{A, D\}, \{B, C\}$

4. $\{A, C\}, \{B, D\}$

Exercise 25 (mult. choice): Consider the set of $N = 4$ points in the two-dimensional space (plane), depicted in the figure. In this case, the dimension of the data space is $l = 2$, so an $(l-1)$-dimensional hyperplane is a straight line in the plane (data space). There are $2^4 = 16$ ways to group the points into two disjoint groups. Which of the following groupings can be implemented via a straight line in the plane (in other words, given a two-group grouping of the data points, is there a line that leaves on its one side the points of the first group and on its other side the points of the second group)?



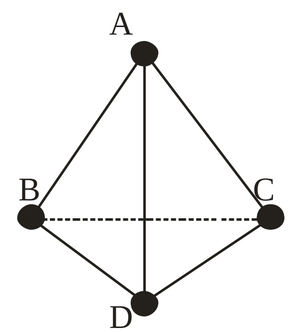
1. $\{A, B\}, \{C, D\}$

2. $\{A\}, \{B, C, D\}$

3. $\{A, D\}, \{B, C\}$

4. $\{A, C\}, \{B, D\}$

Exercise 26 (mult. choice): Consider the set of $N = 4$ points in the three-dimensional space, depicted in the figure (for illustration purposes, the four points are shown as the vertices of a pyramid). In this case, the dimension of the data space is $l = 3$, so an $(l - 1)$-dimensional hyperplane is a plane. Note that in this case, $N = l + 1$. There are $2^4 = 16$ ways to group the points into two disjoint groups. Which of the following groupings can be implemented by a plane in the space (in other words, given a two-group grouping of the data points, is there a plane that leaves on its one side the points of the first group and on its other side the points of the second group)?



1. $\{A, B\}, \{C, D\}$

2. $\{A\}, \{B, C, D\}$

3. $\{A, D\}, \{B, C\}$

4. $\{A, C\}, \{B, D\}$

Exercise 27 (mult. choice): Consider three sets D_1, D_2 and D_3, each one of them consisting of $N = 4$ points. The points of D_1, D_2 and D_3 lie in one, two and three dimensional spaces, respectively. For each one of these data sets, there are $2^4 = 16$ ways to group the corresponding points into two disjoint groups. How many groupings can be implemented by an $(l - 1)$-dimensional hyperplane for each one of the data sets (in other words, given a two-group grouping of the data points, is there an $(l - 1)$-dimensional hyperplane that leaves on its one side the points of the first group and on its other side the points of the second group)?

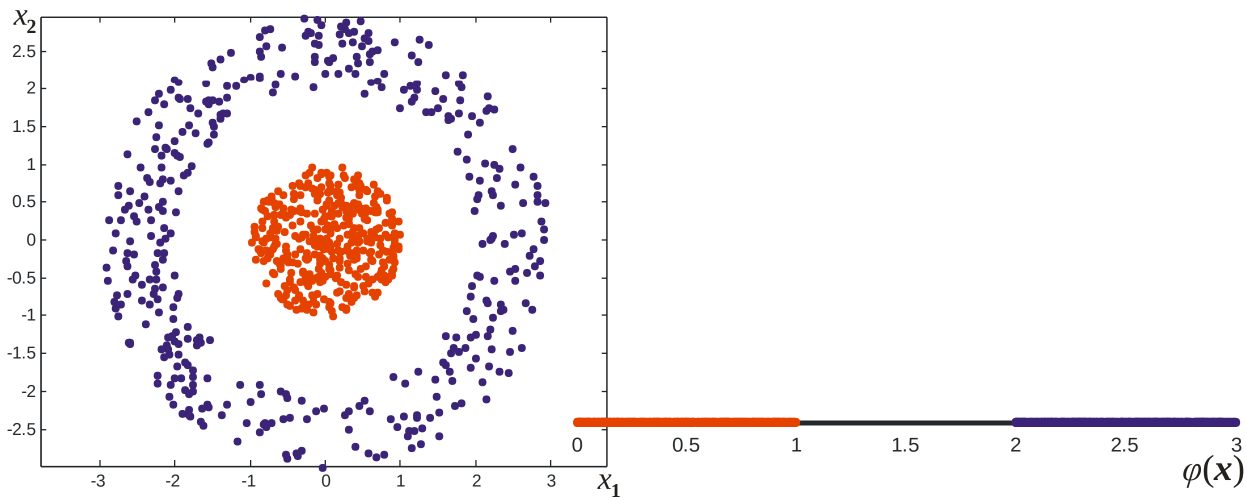
1. D_1: 16, D_2: 14, D_3: 8

2. D_1: 8, D_2: 14, D_3: 16

3. $D_1: 8$, $D_2: 8$, $D_3: 8$

4. $D_1: 14$, $D_2: 14$, $D_3: 14$

Exercise 28 (mult. choice): Consider the two-dimensional two-class classification task shown in the figure (a) below. The (red-colored) points of class $+1$ lie inside the unit radius circle centered at the origin, while the (blue-colored) points from class -1 lie within the ring bordered by two circles centered at the origin and having radius equal to 2 and 3, respectively. Clearly, the two involved classes are not linearly separable. However, if one uses the mapping implied by the function $\varphi(\boldsymbol{x}) = \sqrt{{x_1}^2 + {x_2}^2}$, the original two-dimensional space is mapped to the one-dimensional space, as shown in figure (b), where the two classes become linearly separable. This seems to be in contradiction with the Cover's theorem. The explanation(s) of this behavior is (are):



(a)

(b)

1. Cover's theorem holds only for cases where the dimensionality of the original space is greater than two.

2. Only in the limit $l \to \infty$, linear separability is guaranteed (with high probability).

3. Cover's theorem covers all the cases where the data points are in a general (random) position in the original space, irrespective of whether the points of the two sets are linearly separated, or may exhibit specific structure or not. In this case, the data do not exhibit a random arrangement in the plane.

4. One is allowed to use the mapping function $\boldsymbol{\varphi}(\boldsymbol{x}) = \left[x_1,\ x_2,\ \sqrt{{x_1}^2 + {x_2}^2}\right]^T$, which maps the original space to a three-dimensional one.

Exercise 29 (mult. choice): Consider an l-dimensional two-class classification task, for which a finite set of N data points is available. The two classes are not linearly separable. Assume a mapping that maps an l-dimensional point, $\boldsymbol{x}$, to a K-dimensional space, via the vector function, $\boldsymbol{\varphi}(\boldsymbol{x}) = [\varphi_1(\boldsymbol{x}), \dots, \varphi_K(\boldsymbol{x})]^T$. Consider the two equations, $f(\boldsymbol{x}) = \sum_{k=1}^{K} \theta_k \varphi_k(\boldsymbol{x}) + \theta_0$ and $\hat{y} = \mathrm{sgn}(f(\boldsymbol{x}))$, where $\mathrm{sgn}(z) = 1$, if $z > 0$, and 0, otherwise. Then,

1. These equations define a linear classifier in the transformed space.

2. These equations define a linear classifier in the original space.

3. $f(\boldsymbol{x})$ is linear with respect to the classifier parameters.

4. $\hat{y}$ is linear with respect to the classifier parameters.

Exercise 30 (mult. choice): Consider a two-dimensional two-class classification task where the points from class +1 lie within a circle of unit radius, centered at the origin, while the points from class -1 lie within two circles of unit radius, centered at the points $(10,10)$ and $(-10,-10)$. Clearly, the problem in the two-dimensional space is nonlinearly separable. Which of the following transformations in a higher dimensional space make the problem linearly separable?

1. $\boldsymbol{x} \equiv [x_1,\ x_2]^T \to \boldsymbol{\varphi}(\boldsymbol{x}) = [x_1,\ x_2,\ x_1 +\ x_2\ ]^T$

2. $\boldsymbol{x} \equiv [x_1,\ x_2]^T \to \boldsymbol{\varphi}(\boldsymbol{x}) = \left[x_1,\ x_2, \exp(-\frac{(x_1-10)^2+(x_2-10)^2}{4}) + \exp(-\frac{(x_1+10)^2+(x_2+10)^2}{4})\right]^T$

3. $\boldsymbol{x} \equiv [x_1,\ x_2]^T \to \boldsymbol{\varphi}(\boldsymbol{x}) = \left[x_1,\ x_2, \exp(-\frac{(x_1-x_2)^2}{4})\right]^T$

4. $\boldsymbol{x} \equiv [x_1,\ x_2]^T \to \boldsymbol{\varphi}(\boldsymbol{x}) = \left[x_1,\ x_2, \exp\left(-\frac{{x_1}^2+{x_2}^2}{4}\right)\right]^T$

(D) Nonlinear SVM

Exercise 31 (mult. choice): Consider the following mapping from the two-dimensional space to a four-dimensional space

$$R^2 \ni \boldsymbol{x} = [x_1, x_2] \to \boldsymbol{\varphi}(\boldsymbol{x}) = \left[{x_1}^3, \sqrt{3}{x_1}^2 x_2, \sqrt{3}x_1 {x_2}^2, {x_2}^3\right] \in R^4$$

The inner product of the images of two vectors $\boldsymbol{x}, \boldsymbol{y} \in R^2$ in R^4, $\boldsymbol{\varphi}(\boldsymbol{x})^T \boldsymbol{\varphi}(\boldsymbol{y})$ equals to:

1. $\boldsymbol{x}^T \boldsymbol{y}$

2. $(\boldsymbol{x}^T \boldsymbol{y})^3$

3. $(\boldsymbol{x}^T \boldsymbol{y})^2$

4. $\boldsymbol{x}^T \boldsymbol{y} + \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$

Exercise 32 (mult. choice): Consider the following mapping from the two-dimensional space to the four-dimensional space

$$R^2 \ni \boldsymbol{x} = [x_1, x_2] \to \varphi(\boldsymbol{x}) = \left[{x_1}^3, \sqrt{3}{x_1}^2 x_2, \sqrt{3}x_1 {x_2}^2, {x_2}^3\right] \in R^4.$$

The Euclidean norm of the image of a vector $\boldsymbol{x} \in R^2$ in R^4, $\|\varphi(\boldsymbol{x})\|$, equals to:

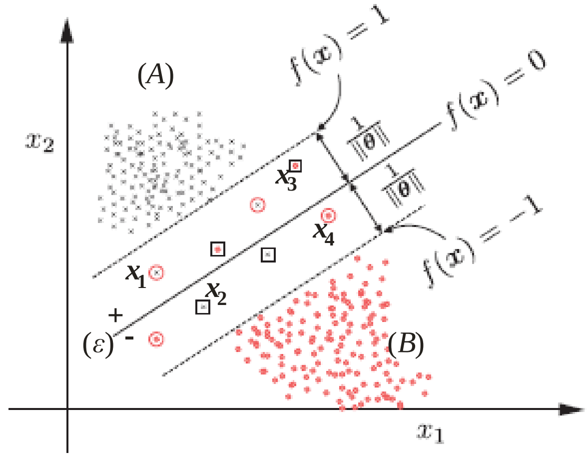
1. $\|\boldsymbol{x}\|^{\frac{2}{3}}$

2. $\|\boldsymbol{x}\|^2$

3. $\|\boldsymbol{x}\|^3$

4. $\|\boldsymbol{x}\|^{\frac{3}{2}}$

Exercise 33 (mult. choice): Consider the two-dimensional two-class classification task shown in the figure on the right, where the graph of $f(\boldsymbol{x}) = \boldsymbol{\theta}^T\boldsymbol{x} + \theta_0 = 0$ is denoted as the solid line (ε) and $f(\boldsymbol{x}) = 1$ and $f(\boldsymbol{x}) = -1$ correspond to the two (dashed) lines (parallel to (ε)) that define the associated margin. Recall that the direction of all lines is determined by the vector $\boldsymbol{\theta}$. The black-colored (resp. red-colored) points stem from class $+1$ (resp. -1). In addition, the points denoted by circles (resp. squares) lie within the margin and are correctly (resp. incorrectly) classified. Which of the following statements is/are true?



1. The points in region (A) are all correctly classified and lie within the margin.

2. The points in region (B) lie outside the margin on the negative side of (ε).

3. Both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ lie outside the margin associated with (ε).

4. The classifier corresponding to (ε) classifies correctly the point $\boldsymbol{x}_4$ and incorrectly the point $\boldsymbol{x}_3$.

(E) Perceptron

Exercise 34 (mult. choice): Consider a two-class (ω_1, ω_2) classification task, for which a finite set of samples $X = \{(y_n, \boldsymbol{x}_n), n = 1, \dots, N\}$, with $y_n \in \{-1, +1\}$, is given ($\omega_1$and ω_2 correspond to $+1$ and -1, respectively). The classification task is linearly separable

1. if there exists a hyperplane that leaves all the vectors $\boldsymbol{x}_n, N = 1, \dots, N$, on its positive side.

2. if there exists a single hyperplane that leaves all y_n's associated with the class ω_1(resp. ω_2) on its positive (resp. negative) side.

3. if there exists a hyperplane that leaves all the vectors, $\boldsymbol{x}_n$, that originate from class ω_1(resp. ω_2) on its positive (resp. negative) side.

4. if only a nonlinear surface can leave the vectors $\boldsymbol{x}_n$ that originate from class ω_1(resp. ω_2) on its positive (resp. negative) side.

Exercise 35 (mult. choice): Consider a two-class (ω_2, ω_1) classification task in the l-dimensional space, for which a finite set of samples $X = \{(y_n, \boldsymbol{x}_n), n = 1, \ldots, N\}$, with $y_n \in \{-1, +1\}$, is given ($\omega_1$and ω_2 correspond to $+1$ and -1, respectively). The involved classes are linearly separable. Which of the following statements is/are true?

1. If $l = 1$ (the feature space is a line), then the separating hyperplane is 1-dimensional (line).

2. If $l = 1$ (the feature space is a line), then the separating hyperplane is a point.

3. If $l = 2$ (the feature space is a line), then the separating hyperplane is 2-dimensional (ordinary plane).

4. If $l = 2$ (the feature space is a line), then the separating hyperplane is 1-dimensional (line).

5. If $l = 3$ (the feature space is a line), then the separating hyperplane is 1-dimensional (line).

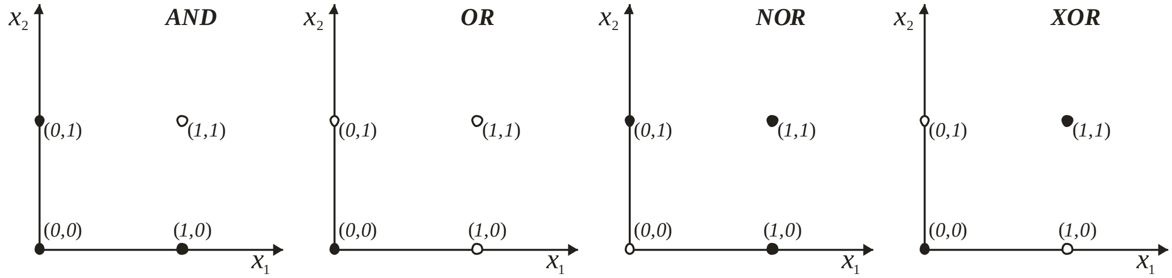
6. If $l = 3$ (the feature space is a line), then the separating hyperplane is 2-dimensional (ordinary plane).

7. If $l = 4$ (the feature space is a line), then the separating hyperplane is 3-dimensional.

8. If $l = 4$ (the feature space is a line), then the separating hyperplane is 4-dimensional.

(F) Feedforward NNs

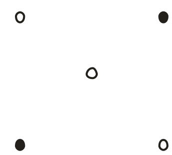
Exercise 36 (mult. choice): Four two-class, two-dimensional classification problems are depicted in the following figures. The points from class ω_1 (resp. ω_2) appear as small open (resp. filled) circles. Which of the following problems **cannot** be solved by a single neuron?



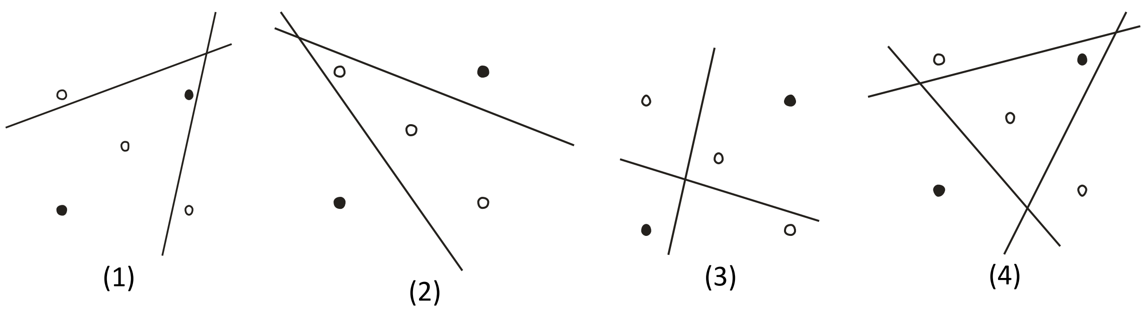
(1) (2) (3) (4)

Hint: Check which of them are linearly separable.

Exercise 37 (mult. choice): Consider the following two-class two-dimensional classification problem where the points from class ω_1 (resp. ω_2) appear as small open (resp. filled) circles.



In which of the following cases the drawn lines define polyhedral regions, i.e., regions whose border is formed by intersecting straight lines , and each of them contains points from a single class?



Exercise 38 (mult. choice): Which of the following statements hold true, with respect to multilayer feedforward neural networks (FNNs), when they are used to solve classification problems where the class regions are unions of polyhedral regions?

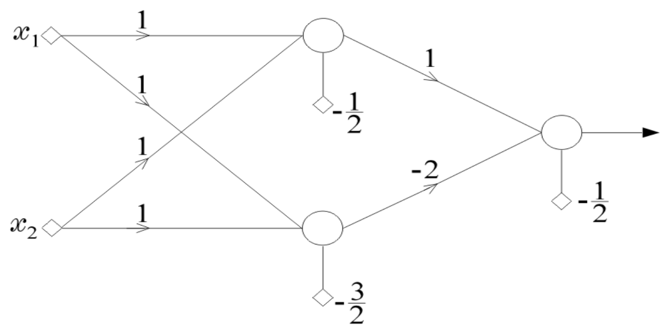
1. Multilayer FNNs cannot handle nonlinearly separable classification tasks.

2. Multilayer FNNs transform gradually a nonlinearly separable classification task to a linear one.

3. All the points from the same polyhedral region have the same representation after the mapping applied by the first layer.

4. Two-layered FNNs are sufficient for solving any classification problem.

Exercise 39 (mult. choice): Consider the following two-layer feedforward neural network (FNN), where each neuron has the Heaviside as the activation function. The FNN solves a two-dimensional two-class classification problem. Which of the following statements is/are true?



1. The FNN classifies $\boldsymbol{x}_1 = [0,0]^T$ to class "1".

2. The FNN classifies $\boldsymbol{x}_1 = [0,0]^T$ to class "0".

3. The FNN classifies $\boldsymbol{x}_2 = [0,1]^T$ to class "0".

4. The FNN classifies $\boldsymbol{x}_2 = [0,1]^T$ to class "1".

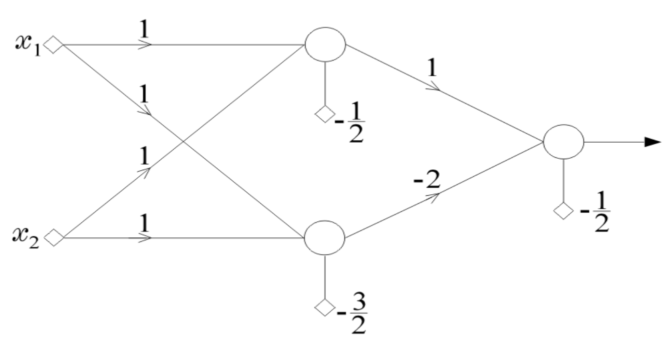
5. The FNN classifies $\boldsymbol{x}_3 = [1,0]^T$ to class "0".

6. The FNN classifies $\boldsymbol{x}_3 = [1,0]^T$ to class "1".

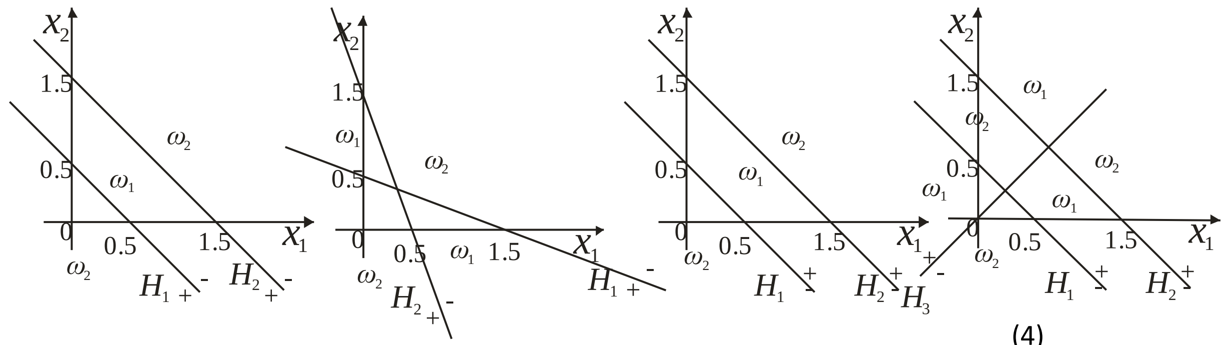
7. The FNN classifies $\boldsymbol{x}_4 = [1,1]^T$ to class "0".

8. The FNN classifies $\boldsymbol{x}_4 = [1,1]^T$ to class "1".

Exercise 40 (mult. choice): Consider the following two-layer feedforward neural network (FNN), where each neuron has the Heaviside as the activation function. The FNN solves a two-dimensional two-class classification problem ($\omega_1(1)$ and $\omega_2(0)$), where the class regions consist of unions of polyhedral regions.

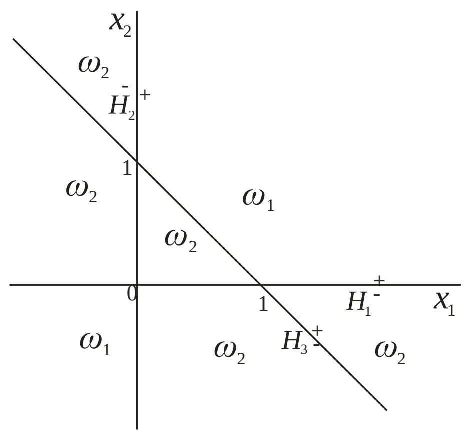


Which of the following figures shows the class regions associated with the above FNN?



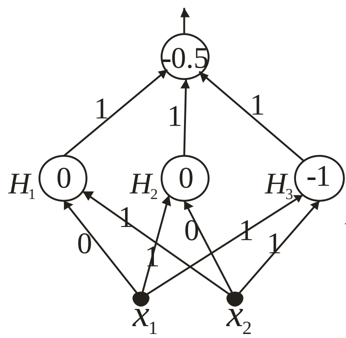
Hint: In order to determine the polarity of a hyperplane (H), choose an arbitrary point $\boldsymbol{x}$ outside (H) and check whether the result obtained after the substitution of the coordinate values of $\boldsymbol{x}$ in the equation of (H) is positive or negative. If it is positive (resp. negative) then the half-space defined by (H) that contains $\boldsymbol{x}$ is the positive (resp. negative) one.

Exercise 41 (mult. choice): Consider the two-class two-dimensional classification problem defined by the following figure.

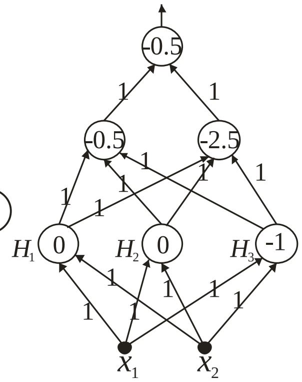


The involved classes are $\omega_1(1)$ and $\omega_2(0)$. The equations of the hyperplanes that form the polyhedral regions that correspond to either of the two classes are:

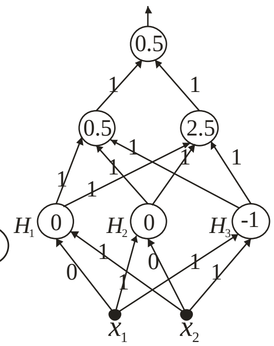
(H_1): $x_2 = 0$, (H_2): $x_1 = 0$ and (H_3): $x_1 + x_2 - 1 = 0$. Which of the following multilayer feedforward neural networks (FNNs), with the Heaviside activation function for all its nodes, solves the classification problem at hand (the bias terms of each neuron is written in its corresponding circle)?



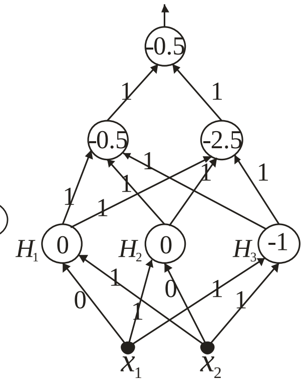
(1)



(2)



(3)



(4)

Exercise 42:

Wolfe dual representation: A convex programming problem is equivalent to

$$max_{\boldsymbol{\lambda}\geq\mathbf{0}} L(\boldsymbol{\theta}, \boldsymbol{\lambda})$$

$$\text{subject to } \frac{\partial}{\partial\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathbf{0}$$

Consider the SVM problem as it is stated in slide 9 of the 9th lecture. Prove that its equivalent dual representation is the one shown in slide 10 of the same lecture.

Hints: (a) The parameters in SVM are $\boldsymbol{\theta}$ and θ_0. Using the Karush-Kuhn-Tacker (KKT) conditions (1) and (2), derive the equations given at the beginning of the 9th slide.

(b) Replace your findings to the Lagrangian function given in the 9th slide of the 9th lecture and perform operations.

(c) Use the Wolfe dual representation given above to state the dual form of the SVM problem (see also the 8th slide).

Exercise 43:

Consider the two-class two-dim. problem where class ω_1 (+1) consists of the vectors $\boldsymbol{x}_1$= [−1, 1]^T, $\boldsymbol{x}_2$= [−1, −1]^T, while class ω_2 (−1) consists of the vectors $\boldsymbol{x}_3$= [1, −1]^T, $\boldsymbol{x}_4$= [1, 1]^T.

(a) **Draw** the points and make a conjecture about the line the (linear) SVM classifier will return.
(b) **Using** the dual representation of the SVM problem, from ex. 42(c) above, derive
   (i) the Lagrange multipliers corresponding to the data points and
   (ii) the line that separates the data points of the two classes.
(c) **Discuss** on the results.

Hints: 1. Define y_1=+1, y_2=+1, y_3=-1, y_4=-1, and substitute to the function

$$(\sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{ij}\lambda_i\lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j) \equiv J_1^*(\boldsymbol{\lambda})$$

the y_i's along with their associated $\mathbf{x}_i$'s. In this way, $J^*_1(\lambda)$ is expressed only in terms of λ_i's (keep in mind that the quantities $\boldsymbol{x}_i^T\boldsymbol{x}_j$ are scalars).

2. Take the derivative of $J^*_1(\lambda)$ with respect to each one of the four λ_i's and set each one of them to zero. Solve the resulting system of equations for λ_i's and find ALL its solutions.

3. Determine the $\boldsymbol{\theta}$ vector, using the equations given in slide 10 of Lecture 9.

4. Determine the θ_0 parameter, by utilizing cond. (4) (see slide 9 of the 9th lecture), for a support vector (a vector whose Lagrange multiplier is positive).

Exercise 44 (Python code + text):

Consider a two-class, two-dimensional classification problem for which you can find attached two sets: one for training and one for testing (file *HW9a.mat*). Each of these sets consists of pairs of the form $(y_i,\boldsymbol{x}_i)$, where y_i is the class label for vector $\boldsymbol{x}_i$. Let N_{train} and N_{test} denote the number of training and test sets, respectively. The data are given via the following arrays/matrices:

- ***train_x*** (a N_{train} x2 **matrix** that contains in its rows the training vectors $\boldsymbol{x}_i$)
- ***train_y*** (a N_{train}–dim. column **vector** containing the **class labels** (0 or 1) of the corresponding training vectors $\boldsymbol{x}_i$ included in ***train_x***).
- ***test_x*** (a N_{test}x2 **matrix** that contains in its rows the test vectors $\boldsymbol{x}_i$)
- ***test_y*** (a N_{test}–dim. column **vector** containing the **class labels** (0 or 1) of the corresponding test vectors $\boldsymbol{x}_i$ included in ***test_x***).

Train the SVM classifier using the training set given above and **measure** its performance using the test set, **using**: (a) the linear kernel, (b) the polynomial kernel and (c) the rbf kernel. Perform **several runs** using the attached code, for several choices of the **parameters** included in each kernel and for various values of C, and comment on the results.