Q1-1: Select the correct statement.

A. *Support vector machines are able to produce non-linear decision boundaries by, in a sense, transforming low-dimensional inputs into a high-dimensional space, then performing classification in that high-dimensional space. This usually works because high-dimensional data is much more likely to be linearly separable than low-dimensional data.*

B. *"Kernel trick" refers to first applying the above transformation and then computing the dot products between the transformed data points.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Q1-1: Select the correct statement.

A. *Support vector machines are able to produce non-linear decision boundaries by, in a sense, transforming low-dimensional inputs into a high-dimensional space, then performing classification in that high-dimensional space. This usually works because high-dimensional data is much more likely to be linearly separable than low-dimensional data.*

B. *"Kernel trick" refers to first applying the above transformation and then computing the dot products between the transformed data points.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Since the training and classification only require the dot products between data points, it "cheats" by using the Kernel function instead. This cheating is referred to as the "kernel trick".

Q1-2: Consider the polynomial kernel $k(x', x') = (xx' + 1)^3$, for $x \in \mathbf{R}$ (i.e., a one-dimensional feature space). Give an explicit expression for the corresponding feature map $\phi(x)$ such that $k(x, x') = \phi(x)^T \phi(x')$.

1. $\phi(x)^T = [x^3, \sqrt{3}\, x^2, \sqrt{3}\, x, 1]$
2. $\phi(x)^T = [x^3, \sqrt[3]{3}\, x^2, \sqrt[3]{3}\, x, 1]$
3. $\phi(x)^T = [x^3, x^2, x, 1]$
4. $\phi(x)^T = [x^3, \sqrt{3}\, x^2, \sqrt{3}\, x]$

Q1-2: Consider the polynomial kernel k(x', x') = $(xx' + 1)^3$, for $x \in \mathbf{R}$ (i.e., a one-dimensional feature space). Give an explicit expression for the corresponding feature map $\phi(x)$ such that $k(x, x') = \phi(x)^T \phi(x')$.

1. $\phi(x)^T = [x^3, \sqrt{3}\, x^2, \sqrt{3}\, x, 1]$
2. $\phi(x)^T = [x^3, \sqrt[3]{3}\, x^2, \sqrt[3]{3}\, x, 1]$
3. $\phi(x)^T = [x^3, x^2, x, 1]$
4. $\phi(x)^T = [x^3, \sqrt{3}\, x^2, \sqrt{3}\, x]$

$$\begin{aligned} k(x, x') &= (xx' + 1)^3 \\ &= (xx')^3 + 3(xx')^2 + 3xx' + 1 \\ &= \begin{bmatrix} x^3 & \sqrt{3}x^2 & \sqrt{3}x & 1 \end{bmatrix} \begin{bmatrix} (x')^3 \\ \sqrt{3}(x')^2 \\ \sqrt{3}x' \\ 1 \end{bmatrix} \end{aligned}$$

Q3-1: Which of the following might be valid reasons for preferring an SVM over a neural network?

A. *An SVM can effectively map the data to an infinite-dimensional space, a neural net cannot.*
B. *The transformed (basis function) representation constructed by an SVM is usually easier to visualize/interpret than for a neural net.*
C. *An SVM would not get stuck in local minima, unlike a neural net.*

1. A, B
2. B, C
3. A, C
4. A, B, C

Q3-1: Which of the following might be valid reasons for preferring an SVM over a neural network?

A. An SVM can effectively map the data to an infinite-dimensional space, a neural net cannot.

B. The transformed (basis function) representation constructed by an SVM is usually easier to visualize/interpret than for a neural net.

C. An SVM would not get stuck in local minima, unlike a neural net.

1. A, B
2. B, C
3. A, C
4. A, B, C

A: True using RBF kernel

B: Not necessarily

C: True (convex optimization problem in SVM)

Q1-1: Are these statements true or false?
(A) If we have multiple optimal solutions on a given training set, those solutions will also have the same test loss.
(B) If a hyperplane only changes its bias term by 1, then the distance from some point x to the hyperplane will not change.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-1: Are these statements true or false?
(A) If we have multiple optimal solutions on a given training set, those solutions will also have the same test loss.
(B) If a hyperplane only changes its bias term by 1, then the distance from some point x to the hyperplane will not change.

1. True, True
2. True, False
3. False, True
4. False, False

(A) Multiple optimal solutions on the training usually have different test loss. Please refer to the example given in the lecture.
(B) Recall that the distance is given by $\frac{|f_{w,b}(x)|}{\|w\|}$. If only the bias term is changed, then $|f_{w,b}(x)|$ will change while $\| w \|$ remains same. So the distance will also be changed.

Q1-2: Are these statements true or false?

(A) Define the margin to be $\gamma = \min_i \frac{y_i f_{w,b}(x_i)}{\|w\|}$, if $f_{w,b}(\mathrm{x})$ predicts correctly on some x_i and incorrectly on others, then the margin will be positive.

(B) If the training set can be correctly separated, then $\max_{w,b} \gamma$ can still be negative.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-2: Are these statements true or false?

(A) Define the margin to be $\gamma = \min_i \frac{y_i f_{w,b}(x_i)}{\|w\|}$, if $f_{w,b}(\mathrm{x})$ predicts correctly on some x_i and incorrectly on others, then the margin will be positive.

(B) If the training set can be correctly separated, then $\max_{w,b} \gamma$ can still be negative.

1. True, True
2. True, False
3. False, True
4. False, False ←

(A) In this case, $\frac{y_i f_{w,b}(x_i)}{\|w\|}$ would be negative on those x_i with incorrect predictions. So take min on all training data, we will get the margin negative.

(B) In this case, there exists at least one w and b such that all instances are correctly classified, so the corresponding margin is non-negative.

Q1-3: Are these statements true or false?
(A) The solution of SVM will always change if we remove some instances from the training set.
(B) If we can only access the labels and the inner products of instances $\{x_i{}^T x_j\}_{i,j}$, we can NOT solve the learning problem in SVM.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-3: Are these statements true or false?
(A) The solution of SVM will always change if we remove some instances from the training set.
(B) If we can only access the labels and the inner products of instances $\{x_i{}^T x_j\}_{i,j}$, we can NOT solve the learning problem in SVM.

1. True, True
2. True, False
3. False, True
4. False, False ⬅

(A) As is shown in the lecture, if we remove those instances with $\alpha_i = 0$, it will not influence the SVM result.
(B) We can see that the dual problem only depends on y_i and the inner products of training instances. So we can also solve the SVM problem in this case.