

GroupE Practice

1. Regarding the SVM:

- (a) For the soft-margin SVM with decision boundary $\mathbf{w}\mathbf{x} + b = 0$ where $\mathbf{w} = (1, 2)$ and $b = 3$, find the hinge loss for $(\mathbf{x}, y) = ((4, 5), -1)$.

Hinge loss is $\max(0, 1 - y_i(\mathbf{w}\mathbf{x} + b)) = \max(0, 1 - (-1)((1 \cdot 4 + 2 \cdot 5) + 3)) = 18$.

- (b) If we train a hard-margin linear SVM on linearly separable data, then discard the support vectors, and then train a new SVM on the remaining examples, the first SVM will have a wider “road” than the second.

☐ True.

☒ False.

- (c) A linear SVM with decision boundary $(1, 2, 2) \cdot \mathbf{x} - 2 = 0$ has a smaller margin between $+1$ and -1 support vectors than one with boundary $(1, 4, 8) \cdot \mathbf{x} + 3 = 0$.

☐ True.

☒ False. The margin for the first SVM is $\frac{2}{\|\mathbf{w}\|} = \frac{2}{\sqrt{1^2 + 2^2 + 2^2}} = \frac{2}{3}$, while the margin for the second is $\frac{2}{\|\mathbf{w}\|} = \frac{2}{\sqrt{1^2 + 4^2 + 8^2}} = \frac{2}{9}$.

2. Consider a database consisting of these three examples:

name $\mathbf{x}$	age y
Leili	20
Layla	22
Kayla	24

We want to estimate Lalla’s age from her name, supposing her name is a corrupted version of one of the names in the database. (It was corrupted, e.g., by a typo.)

- (a) Find the Hamming distance between Lalla and each of the other three names.

name	Hamming distance to Lalla
Leili	ANSWER: 3
Layla	ANSWER: 1
Kayla	ANSWER: 2

- (b) Use 2-NN (two nearest neighbors) regression to estimate Lalla’s age from her name.

Lalla’s age is about years old.

Her 2NN are Layla and Kayla, so we estimate her age as their average age, $\frac{22+24}{2} = 23$.

3. Consider applying gradient descent with step size $\alpha = 0.1$ to find the $\mathbf{x}$ that minimizes the function $f(\mathbf{x}) = f(x^{(1)}, x^{(2)}) = (x^{(1)} + 1)^2 + (x^{(2)} - 2)^2$ starting from $\mathbf{x}_0 = (1, 2)$.

(a) Find the value $\mathbf{x}_1$ after one iteration.

ANSWER:

$\nabla f(\mathbf{x}) = (2(x^{(1)} + 1), 2(x^{(2)} - 2))$, which is $(4, 0)$ at $\mathbf{x} = (1, 2)$.

Move to $\mathbf{x}_1 = \mathbf{x}_0 - \alpha \nabla f(\mathbf{x}_0) = (1, 2) - 0.1(4, 0) = (0.6, 2)$.

- (b) Gradient descent can fail to converge on a convex function if (fixed) step size α is such that we get stuck in a cycle, oscillating between two or several values.

☐ True. ●

☐ False.

- (c) Gradient descent can fail to converge on a convex function if it gets stuck in a local minimum.

☐ True.

☐ False. ●

- (d) Gradient descent can fail to converge on a convex function if the (fixed) step size $\alpha > 0$ is too small.

☐ True.

☐ False. ●

4. Use one-hot encoding to transform the categorical feature `first_name` into binary features with reasonable names that are in alphabetical order.

(input) <code>first_name</code>	(output) ANSWER: <code>Andrew</code> <code>Anna</code> <code>Teresa</code>		
Anna	0	1	0
Teresa	0	0	1
Anna	0	1	0
Andrew	1	0	0