Q1-1: Which of the following statement(s) is(are) TRUE?

A. *Regularization discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.*
B. *Data Augmentation can NOT be considered as a regularization technique.*

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

Q1-1: Which of the following statement(s) is(are) TRUE?

A. *Regularization discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.*

B. *Data Augmentation can NOT be considered as a regularization technique.*

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

Regularization is a technique for combating overfitting and improving training.

Data Augmentation technique generates new training data from given original dataset. It provides a cheap and easy way to increase the amount of your training data.

Q1-2: Which of the following statement(s) is(are) TRUE about regularization parameter λ ?

A. *λ is the tuning parameter that decides how much we want to penalize the flexibility of our model.*
B. *λ is usually set using cross validation.*

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

Q1-2: Which of the following statement(s) is(are) TRUE about regularization parameter λ ?

A. *λ is the tuning parameter that decides how much we want to penalize the flexibility of our model.*
B. *λ is usually set using cross validation.*

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

- The optimization problem can be viewed as following:

$$\text{minimize}(\text{Loss}(\text{Data}|\text{Model}) + \lambda\, \text{complexity}(\text{Model}))$$

- If the regularization parameter is large then it requires a small model complexity
- We have learned how to use cross validate to set hyperparameters including regularization parameters.

Q2-1: Select the correct option about regression with L2 regularization (also called *Ridge Regression*).

A. *Ridge regression technique prevents coefficients from rising too high.*
B. *As $\lambda \rightarrow \infty$, the impact of the penalty grows, and the ridge regression coefficient estimates will approach infinity.*

1. Both statements are true.
2. Both statements are false.
3. Statement A is true, Statement B is false.
4. Statement B is true, Statement A is false.

Q2-1: Select the correct option about regression with L2 regularization (also called *Ridge Regression*).

A. *Ridge regression technique prevents coefficients from rising too high.*

B. *As $\lambda \rightarrow \infty$, the impact of the penalty grows, and the ridge regression coefficient estimates will approach infinity.*

1. Both statements are true.
2. Both statements are false.
3. Statement A is true, Statement B is false. ←
4. Statement B is true, Statement A is false.

As $\lambda \rightarrow \infty$, the impact of the penalty grows, and the ridge regression coefficient estimates will approach zero.

Q2-2: Find the closed-form solution for **w** for the following optimization problem [Ridge Regression].

$$\min_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \ + \ \lambda\|\boldsymbol{w}\|_2^2$$

1. $(XX^T + \lambda I)^{-1}X^Ty$
2. $(X^TX + \lambda I)^{-1}Xy$
3. $(XX^T + \lambda I)^{-1}Xy$
4. $(X^TX + \lambda I)^{-1}X^Ty$

Q2-2: Find the closed-form solution for **w** for the following optimization problem [Ridge Regression].

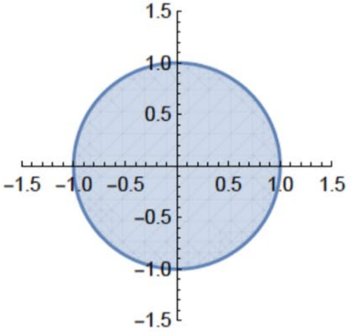
$$\min_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \; + \; \lambda\|\boldsymbol{w}\|_2^2$$

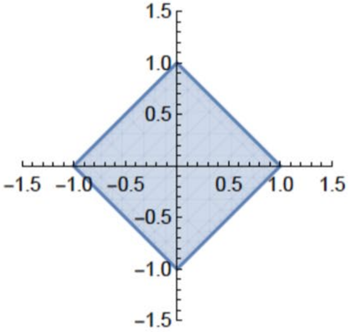
1. $(XX^T + \lambda I)^{-1}X^Ty$
2. $(X^TX + \lambda I)^{-1}Xy$
3. $(XX^T + \lambda I)^{-1}Xy$
4. $(X^TX + \lambda I)^{-1}X^Ty$

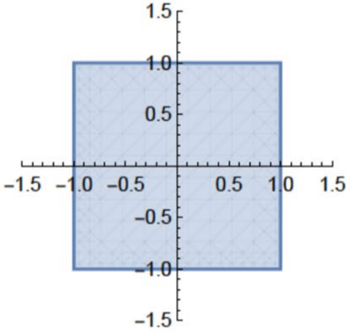
Setting the derivative with respect to **w** to 0 results in:

$$-\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) \; + \; \lambda\boldsymbol{w} \; = \; 0$$

Q3-1: Following figure shows 3-norm sketches: $||x||_p<1$ for p = 1, 2, ∞. Recall that $||x||_\infty = \max\{|x_i| \text{ for all } i\}$

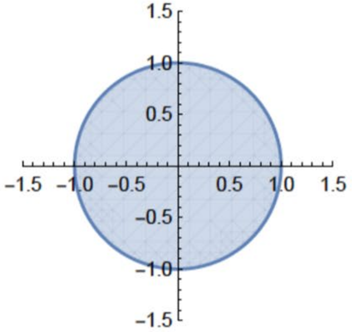
A: 

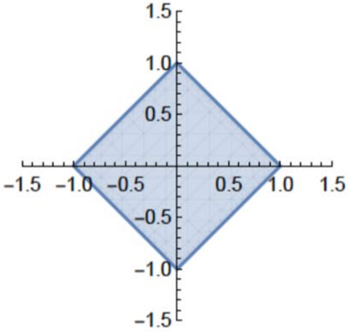
B: 

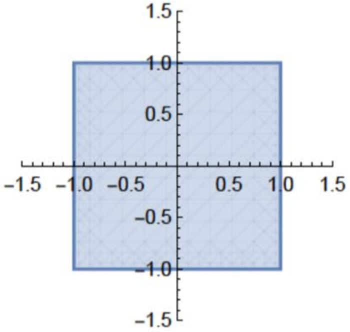
C: 

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

Q3-1: Following figure shows 3-norm sketches: $||x||_p<1$ for p = 1, 2, ∞.
Recall that $||x||_\infty = \max\{|x_i|$ for all i}

A: 

B: 

C: 

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

Q3-2: Find the closed-form solution for **w** for the following optimization problem [LASSO Regression].

$$\min_{\boldsymbol{w}} \left\{ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \; + \; \lambda \|\boldsymbol{w}\|_1 \right\}$$

1. $(X^T X)^{-1}(X^T y - \lambda I)$
2. $(X^T X)^{-1}(X^T y + \lambda I)$
3. $(X^T X)^{-1} X^T y$
4. None of the above

Q3-2: Find the closed-form solution for **w** for the following optimization problem [LASSO Regression].

$$\min_{\boldsymbol{w}} \{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \;+\; \lambda\|\boldsymbol{w}\|_1\}$$

1. $(X^TX)^{-1}(X^Ty - \lambda I)$
2. $(X^TX)^{-1}(X^Ty + \lambda I)$
3. $(X^TX)^{-1}X^Ty$
4. None of the above ←

Setting the derivative with respect to **w** to 0 results in:

$-X^T(y - Xw) + \lambda \, \text{sign}(w) = 0$

$\mathbf{X^TXw + \lambda \, sign(w) = X^Ty}$

No closed form solution exist.