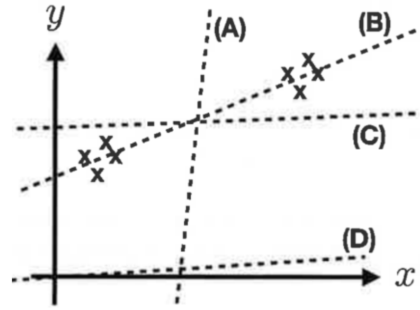
Ahmed Nouralla – B19-CS-01 – a.shaaban@innopolis.university

ML Assignment 1 – Theoretical Part

2.2 Theoretical Question on Ridge Regression

$$(\hat{\theta}, \hat{\theta_0}) = \underset{\theta, \theta_0}{argmin} \sum_{t=1}^{n} (y_t - \theta x_t - \theta_0)^2 + \lambda\theta^2$$



(A) Neither, as ridge regression may never produce a negative $\hat{\theta}_0$ intercept (since both λ and the Residual Sum of Squares RSS cannot be negative).

(B) Low λ (regularization term is small) since the model is almost only affected by the RSS alone with little to no penalization from the L2 regularizer.

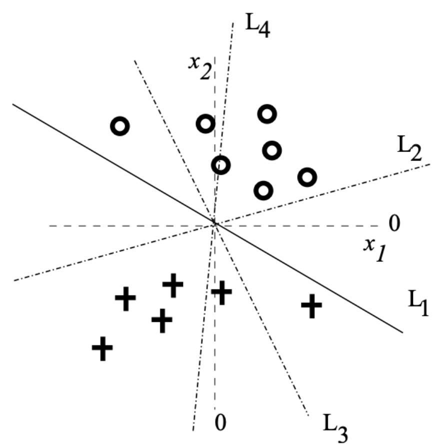
(C) High λ (regularization term is high) since line slope tends to 0 and the model seems over-penalized.

(D) Neither, as regression has to account for the data points. The line depicted has the intercept $\hat{\theta}_0 \cong 0$ which cannot be the argument that minimizes the estimator for the given x, y points.

   a. Ridge regression might only increase intercept and decrease slope than the normal Least Squares line.

3.2 On Regularization in Logistic Regression

$$\sum_{i=1}^{n} log\ p(y_i|x; \theta_1, \theta_2) - \frac{C}{2}\theta_2^2$$



- By increasing the value of parameter C (penalizing θ_2 more), we make the model less sensitive to the changes in x_2
- And thus, the line relies less on the value on x_2 and can only become more vertical.
- Hence, only L_3 may result from regularizing θ_2 as we increase C (which will increase the training error in this case, but may reduce test error, who knows, use cross-validation to tune C)
- L_4 cannot result since regularization can only increase the slope (by penalizing θ_2 more), but not invert its angle.
- If we penalized θ_1 instead of θ_2, L_2 may result in that case.