TP Regularization

We have implemented the logistic regression and linear régrssion in the previous labs, we'll going to improve these algorithms by adding regularization to minimize overfitting and improve the model's ability to generalize.

In this lab and for the two exercises, 80% of the data will be used to train the models, the remaining 20% will be used to test their ability to generalize. Lambda (λ) is a hyperparameter and this determines how severe the penalty is, the regularization parameter needs to be tuned on each model training. In our lab, we only propose to test different values in $[0, +\infty[$ to conclude the impact of this value on the generalization capacity of the model.

1 Regression

In this exercise we use the *California Housing Prices* dataset to train the linear regression model in different scenarios. Compare the following four approaches in terms of generalizability and conclude.

- Simple linear regression

$$J(\beta) = \sum_{i=1}^{m} \left(y_i - x_i\beta\right)^2$$

- Ridge regression

$$J(\beta) = \sum_{i=1}^{m} \left(y_i - x_i\beta\right)^2 + \lambda \sum_{j=1}^{n} \beta_j^2,$$

- Lasso regression

$$J(\beta) = \sum_{i=1}^{m} \left(y_i - x_i\beta\right)^2 + \lambda \sum_{j=1}^{n} \left|\beta_j\right|,$$

- Elastic net

$$J(\beta) = \frac{\sum_{i=1}^{m} \left(y_i - x_i\beta\right)^2}{2m} + \lambda \frac{1-\alpha}{2} \sum_{j=1}^{n} \beta_j^2 + \lambda\alpha \left|\beta_j\right|$$

2 Calassification

This time we will use the *Titanic* dataset, and the logistic regression model to do the binary classification by two approaches, for each approach, the model generalization capacity is calculated, then compare and conclude.

- Simple logistic regression

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left[-y^{(i)} \log\left(h_\theta\left(x^{(i)}\right)\right) - \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right]$$

- Sparse logistic regression

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left[-y^{(i)} \log \left(h_\theta \left(x^{(i)} \right) \right) - \left(1 - y^{(i)} \right) \log \left(1 - h_\theta \left(x^{(i)} \right) \right) \right] + \frac{\lambda}{2m} \sum_{j=1}^{n} |\theta_j|$$

- Ridge logistic regression

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left[-y^{(i)} \log \left(h_\theta \left(x^{(i)} \right) \right) - \left(1 - y^{(i)} \right) \log \left(1 - h_\theta \left(x^{(i)} \right) \right) \right] + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2$$

- Elastic net logistic regression

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left[-y^{(i)} \log \left(h_\theta \left(x^{(i)} \right) \right) - \left(1 - y^{(i)} \right) \log \left(1 - h_\theta \left(x^{(i)} \right) \right) \right] + \lambda \frac{1-\alpha}{2m} \sum_{j=1}^{n} \theta_j^2 + \lambda\alpha \left| \theta_j \right|$$