
Linear Models of Machine Learning

Linear Regression Models

$$y = w[0] * x[0] + w[1] * x[1] + ... + w[p] * x[p] + b$$

Where $x[0]$ to $x[p]$ denotes features of a single datapoint, w and b are parameters of the model that are learned and y is the prediction the model makes.

#TODO

Lasso A lasso regression model restricts coefficients to be close to zero using L1 regularization. As a consequence of this technique some coefficients are exactly zero and these are ignored by the model. This can be seen as a form of automatic feature selection and can result in a model being easier to interpret and highlights the most important features of said model

Linear Classification Models

$$y = w[0] * x[0] + w[1] * x[1] + ... + w[p] * x[p] + b > 0$$

Similar to the linear regression formula but instead of returning the weighted sum of the features the predicted value is threshold at zero, i.e if the result is smaller than zero the class is predicted as -1 and if larger than zero the class is predicted as +1

Linear Support Vector Classification (SVG) and Support Vector Machines (SMG) A linear classification model that by default applies L2 regularization

Linear Models for Multiclass Classification

$$w[0] * x[0] + w[1] * x[1] + ... + w[p] * x[p] + b$$

Where w is one vector of coefficients and b one intercept for each class

Many linear classification models don't extend naturally to multiclass cases (except logistic regression). A technique to extend a binary classification algorithm to a multiclass one is the one-vs-rest approach. In this approach a binary model is learned for each class that tries to separate that class from all the other classes resulting in a binary model for each possible class. The new algorithm then makes a prediction by running all the binary classifiers on a test data point and the classifier with the highest score on its single class wins and its label is returned.

Strengths Weaknesses and Parameters

Parameters

- Regularization Intensity - The main parameter of linear models, denoted by alpha in regression models and C in LinearSVC and Logistic Regression. Large values for alpha or small values for C mean simple models and vice versa. These parameters are usually searched for on a logarithmic scale
- Regularization Type - L1 or L2 regularization. L1 is usually used when only a few features are actually important and can be useful to ensure better interpretability of the model

Strengths

- Fast training and prediction
- Scales easily
- Works well with sparse data
- Prediction process is relatively easy to understand using formulae
- Performs well when number of features is greater than number of samples

Weaknesses

- It is often not entirely clear why coefficients are the way they are
- Coefficients can be difficult to interpret if dataset contains highly correlated features
- May yield worse generalization performance in relation to other models in lower-dimensional spaces