Machine Learning

Regression

Claudio Sartori

DISI
Department of Computer Science and Engineering – University of Bologna, Italy
claudio.sartori@unibo.it

Regression – Forecasting continuous values

- Supervised task
- The target variable is numeric
- Minimize the error of the prediction with respect to the target

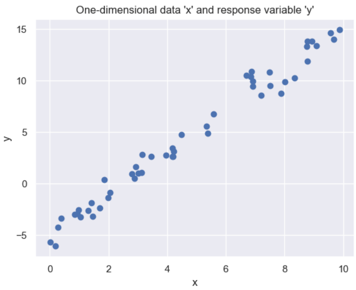
Linear Regression

- data set $\mathcal{X}$ with N rows and D columns
  - x_i is a D dimensional data element
- response vector $\overline{y}$ with N values y_i
- w is a D-dimensional vector of coefficients that needs to be learned
- we model the dependence of each response value y_i from the corresponding independent variables x_i as

$$y_i \approx w^T \cdot x_i \quad \forall i \in [1 \ldots N]$$

- such that the error of modelling is minimised
- Classical statistic method (1805)

Data and regression line



One–dimensional data and response variable

Regression and score - Score range $(-\inf : 1)$

Objective function and minimisation I

$$\mathcal{O} = \sum_{i=1}^{N} (w^T \cdot x_i - y_i)^2 = ||Xw^T - y||^2$$
$$= (Xw^T - y)^T.(Xw^T - y)$$

Gradient of $\mathcal{O}$ with respect to w

$$2X^T(Xw^T - y)$$

Constraining the gradient to 0 we obtain the optimisation condition

$$X^T X w^T = X^T y$$

Objective function and minimisation II

If the symmetric matrix $X^T X$ is *invertible* the solution can be derived as

$$w = (X^T X)^{-1} X^T y$$

and the forecast is given by

$$y^f = X \cdot w^T$$

Matrix calculus

- Issues related to matrix calculus if $\overline{x}^T\overline{x}$ is not invertible
- *Moore–Penrose pseudoinverse*
- *Tikonov regularisation* (also known as *ridge regression*)
- *Lasso regularisation*

Quality of the fitting - R^2

Mean of the observed data	$y^{avg} = \frac{1}{N} \sum_i y_i$
Sum of squared residuals	$SS_{res} = \sum_i (y_i - y_i^f)^2$
Total sum of squares	$SS_{tot} = \sum_i (y_i - y^{avg})^2$

Coefficient of determination $\mathbf{R^2 = 1 - \frac{SS_{res}}{SS_{tot}}}$

Intuition about R^2

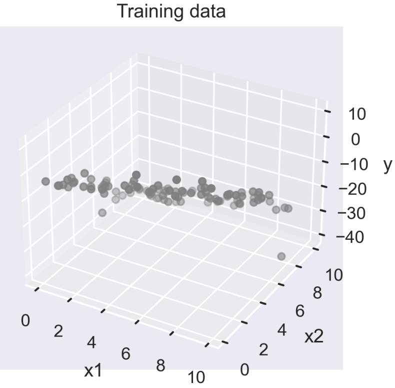
- It compares the fit of the chosen model with that of a horizontal straight line
- With perfect fitting the numerator of the second term is zero and $R^2 = 1$
- If the model does not follow the trend of the data the numerator of the second term can reach or exceed the denominator, and R^2 can also be negative
- Despite the name, R^2 isn't the square of anything

R^2 and *Mean Squared Error*

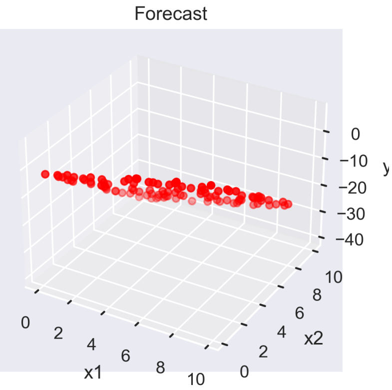
- Both refer to the error of the predictions
- R^2 is a standardised index,
- *RMSE* measures the mean error, this it is influenced by the order of magnitude of the data,
- Both *RMSE* and R^2 quantifies how well a linear regression model fits a dataset
- The RMSE tells how well a regression model can predict the value of a response variable in absolute terms
- R^2 tells how well the predictor variables can *explain the variation in the response variable*
- For comparing the accuracy among different linear regression models, RMSE is a better choice than R Squared
- R^2 is not meaningful for non–linear or non–algebraic regression models

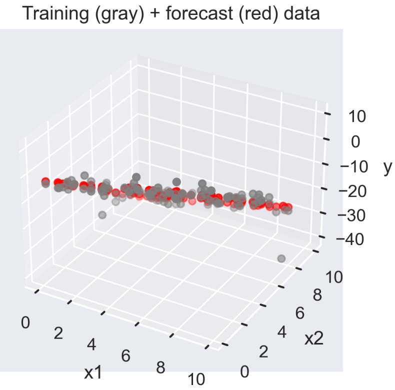
Multiple regression

- The response variable depends by more than one features
- The regression technique is quite similar to that of simple regression
- In `scikit-learn` the estimator is the same



Multiple regression - forecast



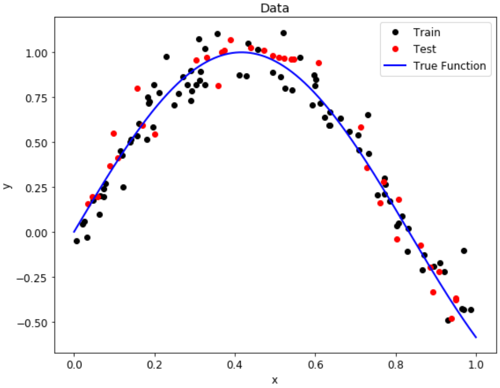


Overfitting and Regularisation

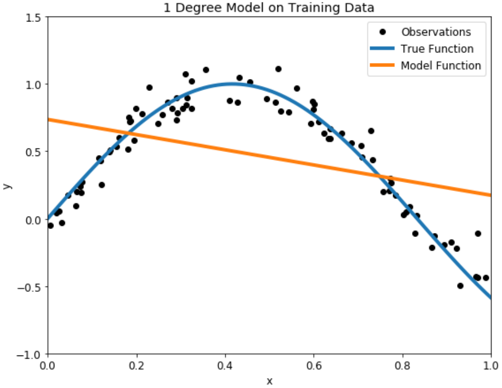
- In presence of high number of features overfitting is possible
  - performance on test data becomes much worse
- Regularisation reduces the influence of less interesting attributes and therefore reduces overfitting

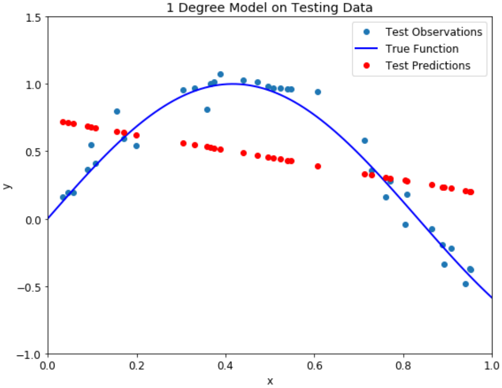
Polynomial regression

- Target is influenced by a single feature
- The relationship cannot be described by a straight line

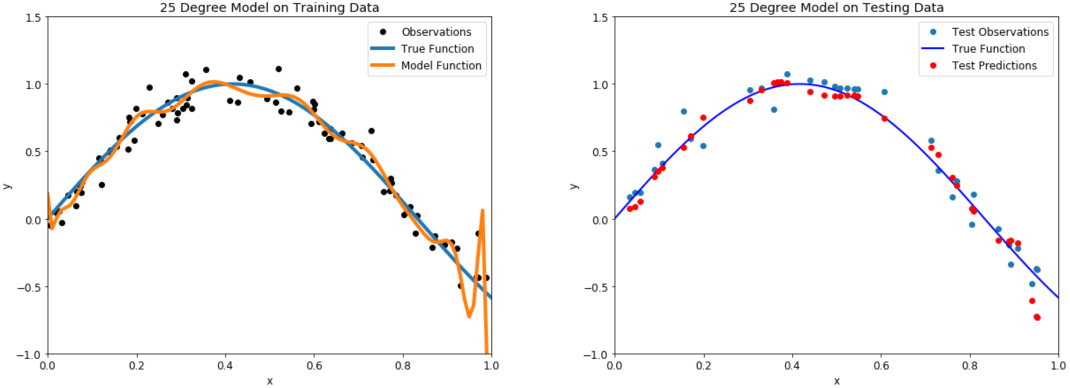


Underfitting

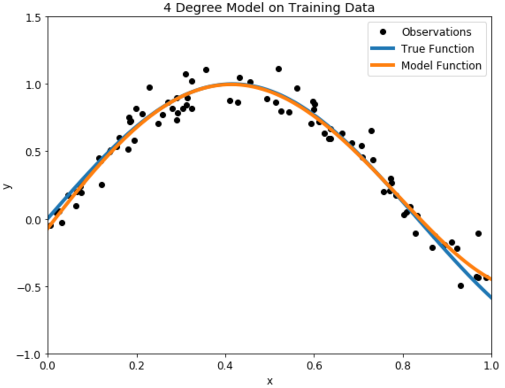


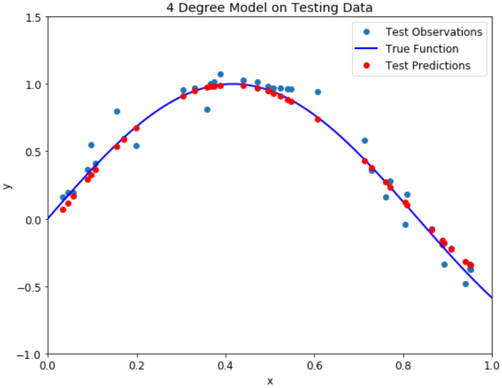


Overfitting

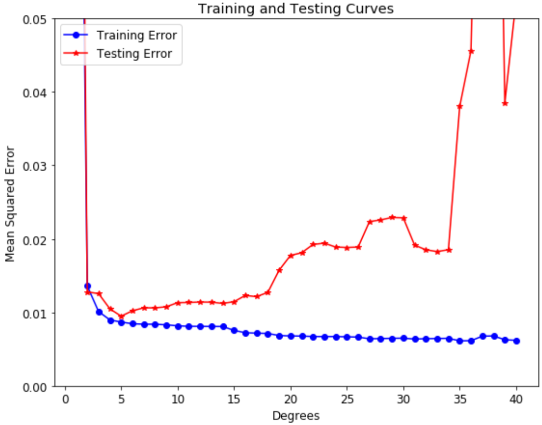


Good fitting





Model complexity vs fitting



Selection of Regression Models

Method	*Library*	*Model Name*
Linear Regression	sklearn.linear_model	LinearRegression
Elastic Net Regression	sklearn.linear_model	ElasticNet
Stochastic Gradient Descent Regression	sklearn.linear_model	SGDRegressor
Bayesian Ridge Regression	sklearn.linear_model	BayesianRidge
Lasso Regression	sklearn.linear_model	Lasso
Support Vector Machine	sklearn.svm	SVR
Kernel Ridge Regression	sklearn.kernel_ridge	KernelRidge
Gradient Boosting Regression	sklearn.ensemble	GradientBoostingRegressor
XGBoost Regressor	xgboost	XGBRegressor
CatBoost Regressor	catboost	CatBoostRegressor
LGBM Regressor	lightgbm	LGBMRegressor