

岭回归器的详细步骤

- 线性回归的主要问题是异常值敏感。在真实世界的的数据收集过程中，经常会遇到错误的度量结果。而线性回归使用的普通最小二乘法，其目标是使平方误差最小化。这时，由于异常值误差的绝对值很大，因此会引起问题，从而破坏整个模型。
- 普通最小二乘法在建模时会考虑每个数据点的影响，因此，最终模型就会偏差很大。显然，我们发现这个模型不是最优的。为了避免这个问题，我们引入正则化项的系数作为阈值来消除异常值的影响，这个方法被称为岭回归。
- **详细步骤**
 - 文件中加载数据。这个文件的每一行都包含多个数值。除了最后一个数值外，前面的所有数值构成输入特征向量。
 - 和线性回归器一样的步,但是构建对象的
 - `ridge_regressor = linear_model.Ridge(alpha=0.01, fit_intercept=True,max_iter=10000)`
 - `alpha`参数控制回归器的复杂程度。当`alpha`趋于0时，岭回归器就是用普通最小二乘法的线性回归器。因此，如果你希望模型对异常值不那么敏感，就需要设置一个较大的`alpha`值。这里把`alpha`值设置为0.01。
 - 训练岭回归器
 - `ridge_regressor.fit(X_train, y_train)`
 - `y_test_pred_ridge = ridge_regressor.predict(X_test)`
 - `print "Mean absolute error =", round(sm.mean_absolute_error(y_test, y_test_pred_ridge), 2)`
 - `print "Mean squared error =", round(sm.mean_squared_error(y_test, y_test_pred_ridge), 2)`
 - `print "Median absolute error =", round(sm.median_absolute_error(y_test, y_test_pred_ridge), 2)`
 - `print "Explain variance score =", round(sm.explained_variance_score(y_test, y_test_pred_ridge), 2)`
 - `print "R2 score =", round(sm.r2_score(y_test, y_test_pred_ridge), 2)`
 - 可以用同样的数据建立一个线性回归器，并与岭回归器的结果进行比较，看看把正则化引入回归模型之后的效果如何。

