Diabetes dataset

The diabetes dataset consists of 10 physiological variables (age, sex, weight, blood pressure) measure on 442 patients, and an indication of disease progression after one year:

```
>>> diabetes = datasets.load_diabetes()
>>> diabetes_X_train = diabetes.data[:-20]
>>> diabetes_X_test  = diabetes.data[-20:]
>>> diabetes_y_train = diabetes.target[:-20]
>>> diabetes_y_test  = diabetes.target[-20:]
```

The task at hand is to predict disease prediction from physiological variables.

Linear Regression

$$\text{Linear models: } \mathbf{y} = \mathbf{X}\beta + \epsilon$$

- X: data
- y: target variable
- β: Coefficients
- ϵ: Observation noise

```
>>> from scikits.learn import linear_model
>>> regr = linear_model.LinearRegression()
>>> regr.fit(diabetes_X_train, diabetes_y_train)
LinearRegression(fit_intercept=True)
>>> print regr.coef_
```

```
[  3.03499549e-01  -2.37639315e+02   5.10530605e+02   3.27736980e+02
  -8.14131709e+02   4.92814588e+02   1.02848452e+02   1.84606489e+02
   7.43519617e+02   7.60951722e+01]

>>> # The mean square error
>>> np.mean((regr.predict(diabetes_X_test) - diabetes_y_test)**2)
2004.5676026898223

>>> # Explained variance score: 1 is perfect prediction
>>> regr.score(diabetes_X_test, diabetes_y_test)
0.58507530226905713
```