REGRESSION MODELS

PERFORMANCE MEASURES

The Learning Penalty

(aka the Penalty Function)

$w_0 x_0 + w_1 x_1 + w_2 x_2$

Penalty

	Feature1	Feature2	Output		
	x_1	x_2	y	$\hat{y}$	$(y-\hat{y})^2$
row 1					
⋮	⋮	⋮	⋮		⋮
row m					

What you deal with in orange

Cost = Sum of all penalties

(add up the value of each row)

More precisely, Cost $= \frac{1}{m}$ (Sum of all penalties)

this makes it the <u>average</u> penalty

How Learning Happens

Iteration 1

usually just randomly chosen values

$w_0 = 0.5, \; w_1 = 2.2, \; w_2 = -1.8$

Penalty $(y-\hat{y})^2$ for each row

Add up the values

Divide by m (# of rows)

↓

Cost of iteration

Goal: Find the next set of w_0, w_1, w_2 that lower this cost.

Gradient descent is the method for doing this.

How Learning Happens (Contd.)

Iteration 2

$w_0 = 0.52$, $w_1 = 2.1$, $w_2 = -1.92$

→ Penalty $(y-\hat{y})^2$ for each row

Add up the values

Divide by m (# of rows)

↓

Cost of iteration

Goal: Find the <u>next set</u> of w_0, w_1, w_2 that **lower** this cost.

↓

Iteration 3, 4, 5, 6, ..., 23

We can choose to stop after a certain # of iterations, or when cost isn't going down by a lot any more.

Summary - For each iteration,

Penalty $= (y-\hat{y})^2$ for each row

$$\text{Cost} = \underbrace{\frac{1}{m} \, \text{Sum}\left[(y-\hat{y})^2\right]}_{\text{Average or Mean Penalty per iteration}}$$

rows in the dataset

Cost = "Mean Squared" Penalty

Once we get to the last iteration, we have the optimal values for w_0, w_1, and w_2.

MSE is the <u>last</u> cost value -
the cost at the last iteration
of the gradient descent algorithm.

Once gradient descent is finished,
you get the optimal parameter values
w_0, w_1, and w_2.

To get the MSE, take every row
of the training dataset and calculate
the penalty. Then add up the
penalties to get the cost.

MSE - Mean Square Error

$$\hat{y} = w_0 x_0 + w_1 x_1 + w_2 x_2$$

Features

x_1	x_2	Prediction	Actual	$(\text{Prediction} - \text{Actual})^2$
				23.42
				5.8

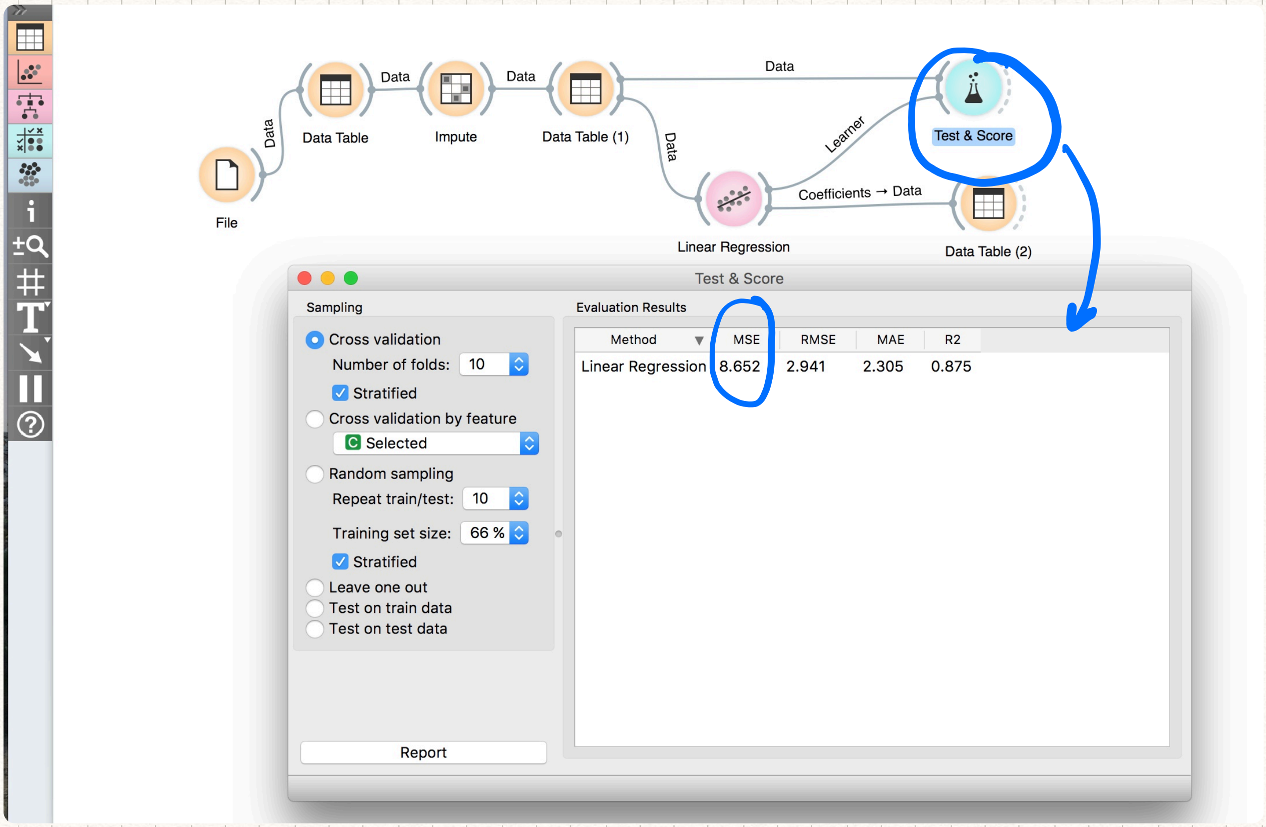
This is the training dataset

$$MSE = \frac{\left[\dfrac{(23.42 + 5.8)}{2}\right]}{\#\text{ of rows}}$$

Identical to the final cost in gradient descent.

Performance

How good is the model ?



MSE = Mean Square Error

RMSE = Root Mean Square Error

MAE = Mean Absolute Error

R2 = R^2 = "R-Squared"

Always a good idea to have a single number (if at all possible) to measure how good your model is.

MSE, RMSE, MAE and R^2 are such numbers. Any <u>one</u> of them can be used to measure how good the model is.

The lower the value, the better for MSE, RMSE, & MAE.

SOME FORMULAS

(OPTIONAL)

SS = Sum of Squares

$\bar{y}$ = average value of actual outputs

$y^{(i)}$ = actual output value of the i^{th} row of the dataset.

$\hat{y}^{(i)}$ = predicted output value of the i^{th} row of the dataset.

$|a-b|$ = Absolute value of $a-b$.

$$SS_{total} = \underset{\text{all rows}}{\text{Sum over}} \left[\left(y^{(i)} - \bar{y} \right)^2 \right]$$

$$SS_{regression} = \underset{\text{all rows}}{\text{Sum over}} \left[\left(\hat{y}^{(i)} - \bar{y} \right)^2 \right]$$

$$\rightarrow SS_{residual} = \underset{\text{all rows}}{\text{Sum over}} \left[\left(y^{(i)} - \hat{y}^{(i)} \right)^2 \right]$$

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}}$$

MSE = Mean Square Error

$$= \frac{SS_{residual}}{\text{\# of rows in dataset}}$$

RMSE = Root Mean Square Error

$$= \sqrt{MSE}$$

MAE = Mean Absolute Error

Absolute value

$$\frac{\text{Sum over all rows}\left[\,|\hat{y}^{(i)} - y^{(i)}|\,\right]}{\text{\# of rows in the dataset}}$$

SUMMARY

- Compare regression models using their RMSE Scores.

- The lower the RMSE, the better the model.