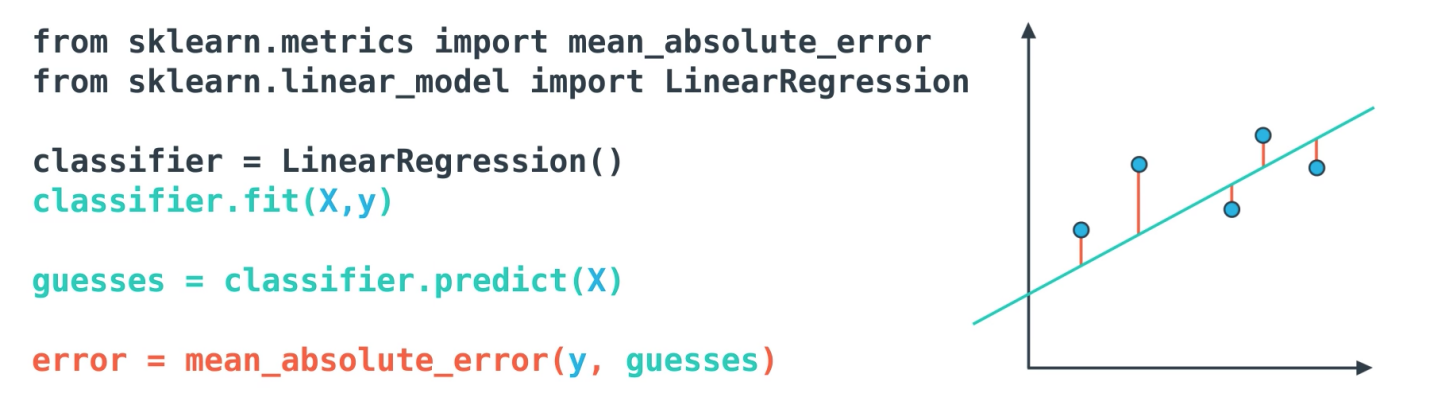
# Mean-Squared Error (MSE) Evaluating regression models

## Mean Absolute Error (MAE)

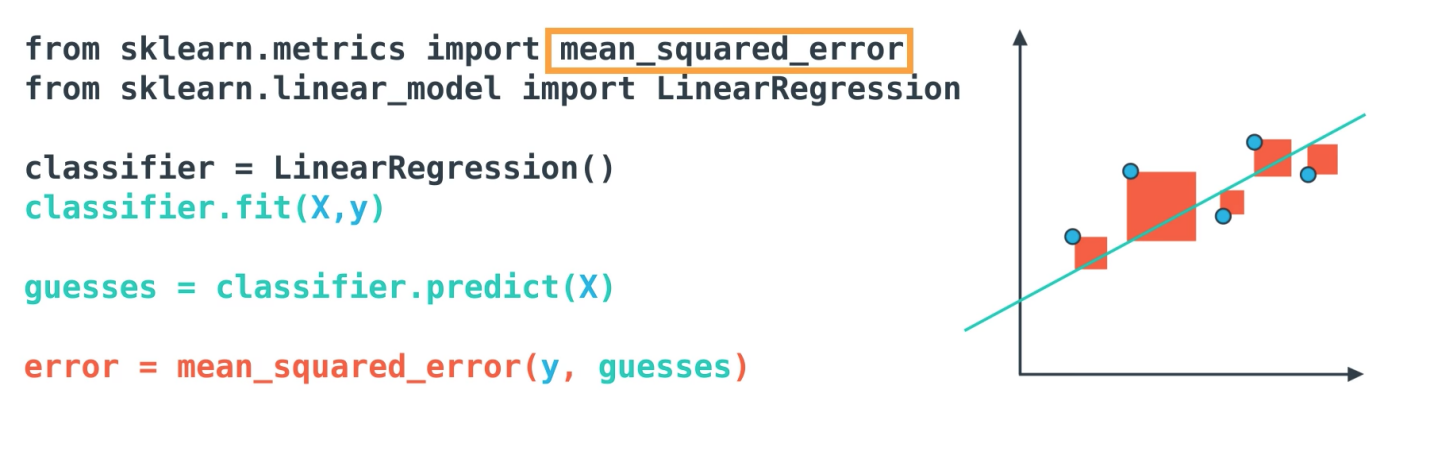
* In this example, we have some points and we have drawn a line that fits these points. Now we would like to check how bad this line is. A good way to do this is to just add the absolute values of the distances from the points to the line. Mean absolute error is very easy to calculate in SKLearn. Create a linear regression classifier object called classifier.
* Then we use the function to fit the line. Now we can refer to the points as y and to the predictions made by the model as classifier.predict(X). We will call these guesses. Finally, the mean absolute error is just calculated with the mean absolute error function. The mean absolute error has a problem which is that the absolute value function is not differentiable. This may not be good if we want to use methods such as gradient descent.
* This is a useful metric to optimize when the value you are trying to predict follows a skewed distribution. Optimizing on an absolute value is particularly helpful in these cases because outliers will not influence models attempting to optimize on this metric as much as if you use the mean squared error. The optimal value for this technique is the median value.



## Mean-Squared Error (MSE)

The mean squared error is by far the most used metric for optimization in regression problems. Similar to MAE, you want to find a model that minimizes this value. This metric can be greatly impacted by skewed distributions and outliers. When a model is considered optimal via MAE, but not for MSE, it is useful to keep this in mind. In many cases, it is easier to actually optimize on MSE, as the quadratic term is differentiable. However, an absolute value is not differentiable. This factor makes this metric better for gradient-based optimization algorithms.

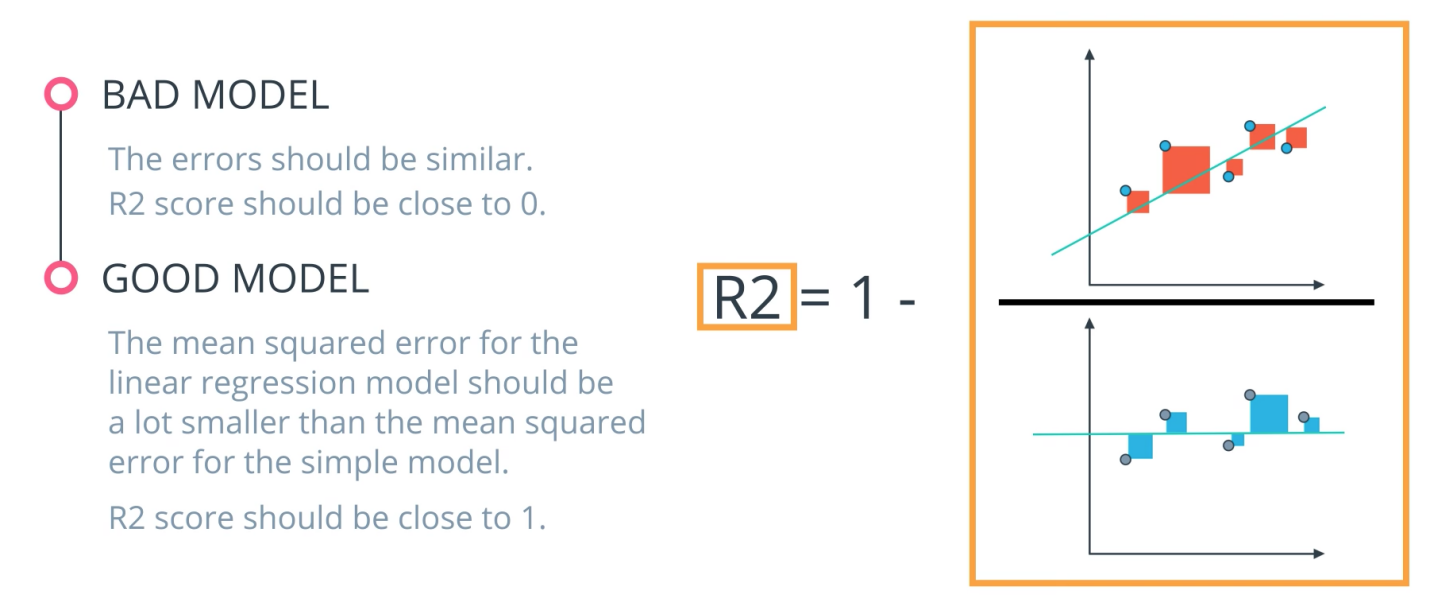
For this metric, we add the squares of the distances between the points and the line. Mean square error is also very easy to calculate in SKLearn, except now we will use the mean square error function.



Mean Squared Error

## R2 score

* Finally, the r2 value is another common metric when looking at regression values. Optimizing a model to have the lowest MSE will also optimize a model to have the highest R2 value. This is a convenient feature of this metric. The R2 value is frequently interpreted as the 'amount of variability captured by a model. Therefore, you can think of MSE, as the average amount you miss across all the points and the R2 value as the amount of the variability in the points that you capture with a model.
* R2 score is based on comparing our model to the simplest possible model. Let's think, what is the simplest possible model that fits a bunch of points? Well, a pretty simple one is just to take the average of all the values and draw a horizontal line through them, and we can calculate the mean squared error for this model. We would hope that the mean squared error for the simple model is larger than the error for a linear regression model. The question is, how much larger? Well, we can divide the error for the linear regression model by the error for the simple model, and then subtract the result from 1, and we will call these the R2 score. If the model is not very good, then the two errors should be similar and this quantity here should be close to one. The whole R2 score should be close to zero. If the model is good, then the mean squared error for the linear regression model should be a lot smaller than the mean square for the simple model. Therefore, this ratio should be small and then the R2 squared should be very close to one.
* As you can see in the image below, if the R2 score is close to one, then the model is good. If it's close to zero, then the model is not much better than just guessing the average of the values of the points R2 score is very simple to calculate in SKLearn with the R2 score function.



R2 Score

## Summary

Do you want to measure how well your algorithms are performing on predicting numeric values? In these cases, there are three main metrics that are frequently used. **mean absolute error**, **mean squared error** and **r2** values.

As an important note, optimizing on the mean absolute error may lead to a different 'best model' than if you optimize on the mean squared error. However, optimizing on the mean squared error will **always** lead to the same 'best' model as if you were to optimize on the **r2** value.

Again, if you choose a model with the best r2 value (the highest), it will also be the model that has the lowest (MSE). Choosing one versus another is based on which one you feel most comfortable explaining to someone else.

### **Quiz Question**

Connect the evaluation metric with the correct description

Divide the error for the linear regression model by the error for the simple model, and then subtract the result from 1

Answer: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Add the squares of the distances between the points and the line

Answer: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Add the absolute values of the distances from the points to the line.

Answer: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Divide the error for the linear regression model by the error for the simple model, and then subtract the result from 1

Answer: R-squared (R²)

2. Add the squares of the distances between the points and the line

Answer: Mean Squared Error (MSE)

3. Add the absolute values of the distances from the points to the line

Answer: Mean Absolute Error (MAE)

Extra Resource: <https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-machine-learning-tips-and-tricks>