Introducing Scikit-Learn

Bookmark this page

We can do the model training using the scikit-learn library. These details are also a recap from the exercise "Linear Regression Using Sklearn".

First we import the linear regression module

```
from sklearn.linear_model import LinearRegression
```

We use pandas and the method `read_csv` to read the advertising data and assign it to a data frame which we call df

```
df = pd.read_csv('Advertising.csv')
```

We take the TV column and the sales column and assign the series to X and y

```
X = df[['TV']].values
y = df['Sales'].values
```

We instantiate the model by using linear regression and call it `reg`

```
reg = LinearRegression()
```

We use now that model to fit the data by passing the data X and y

```
reg.fit(X,y)
```

Once we have fitted the model we can use the same variable, reg, to take the attribute coefficients to see the slope and intercept. The underscore is used to see the slope.

```
reg.coef_
>>> array[[0.04665056]])

reg.intercept_
>>> array[[7.08543108]])
```

We can use the method predict to predict at any given x what would be the value of y

```
reg.predict(np.array([[100]])
>>> array([[11.75048733]])
```

What is happening behind the scenes? This code is hiding a lot of things, and we should investigate what's happening at the `.fit` method in order to understand how we find the β_0 and β_1 that minimizes the loss function.

Calculus Review

Derivative definition

First we have to review some concepts. A derivative is the instantaneous rate of change or slope of a single valued function given. Given a function $f(x)$ the derivative can be defined as the slope as the two points come closer to each other where the difference is 0. As shown in the figure, the derivative is the slope between the blue point and the red point as the two points come closer to each other.

$$f'(x) = \frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Partial derivatives

Another concept is the partial derivative. The partial derivative is for a function that depends on two or more variables. The rate of change of a function with respect to one variable while the other is fixed is called a partial derivative.

For a function f, the partial derivative is written as

$$\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots$$

So, we have a function that depends on x_1 and x_2 the partial derivative of f with respect to x_1 is the slope when x_2 is kept fixed and the partial derivative of f with respect to x_2 is the slope whe

Calculator

fixed.

fixed.

$$f(x_1, x_2) = \begin{cases} \frac{\partial f}{\partial x_1} & \text{for } x_1 \\ \frac{\partial f}{\partial x_2} & \text{for } x_2 \end{cases}$$

Optimization

In our simple linear regression, our loss functions depend on β_0 and β_1. To minimize a loss function, we need to determine the rate of change of the function with respect to one variable with the others held fixed.

The global minima or maxima of the loss function $L(\beta_0, \beta_1)$ must occur at a point where the gradient (slope) is equal to zero, so the partial derivatives here must be equal to zero.

$$\nabla(L) = \left[\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1}\right] = 0$$

There are three ways we can do this:

- **Brute force:** Try every combination of β_0 and β_1 and find where both partial derivatives are equal to 0
- **Exact:** Solving the above equations only works for rare cases. Since there are 2 equations and 2 unknowns in our example, it is solvable.
- **Greedy Algorithm:** Gradient descent will be discussed later

Partial derivative example

For our linear regression, we are going to solve it analytically exactly. We start with the loss function that depends on β_0, β_1 and want the partial derivatives with respect to β_0 and β_1

If $L(\beta_0, \beta_1) = (y - (\beta_1 x + \beta_0))^2$ then what are $\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1}$?

We need the chain rule to solve this. The chain rule says that if a function depends on another function that depends on a variable, the partial derivative of the original function with respect to the variable is the product of two partial derivatives. The first is the partial derivative of the function with the intermediate function multiplied by the partial derivative of the intermediate function with respect to that variable.

We can use that chain rule to derive the partial derivatives of the loss function with respect to β_0 and the partial derivative of the loss function with respect to β_1

If $L(\beta_0, \beta_1) = (y - (\beta_1 x + \beta_0))^2$ then waht is $\frac{\partial L}{\partial \beta}$?

$$\frac{\partial L(f(\beta_0))}{\partial \beta_0} = \frac{\partial L}{\partial f}\frac{\partial f}{\partial \beta_0}$$

Partial Derivative $\frac{\partial L}{\partial \beta_0}$

We can use some algebra

If $L(\beta_0, \beta_1) = (y - (\beta_1 x + \beta_0))^2$ then waht is $\frac{\partial L}{\partial \beta_0}$?

$$L = ((y - \beta_1 x - \beta_0)^2)$$

where $f = (y - \beta_1 x - \beta_0)$ and $L = f^2$

$$\frac{\partial L}{\partial \beta_0} = \frac{\partial L}{\partial f}\frac{\partial f}{\partial \beta_0}$$

$$L = f^2$$

$$\frac{\partial L}{\partial f} = 2f$$

$$f = y - \beta_1 x - \beta_0$$

Calculator

$$\frac{\partial f}{\partial \beta_0} = -1$$

$$\frac{\partial L}{\partial \beta_0} = \frac{\partial L}{\partial f}\frac{\partial f}{\partial \beta_0} = -2f = -2\left(y - \beta_1 x - \beta_0\right)$$

Partial Derivative $\frac{\partial L}{\partial \beta_1}$

We can use the same algebra used for β_0 to find $\frac{\partial L}{\partial \beta_1}$

If $L\left(\beta_0, \beta_1\right) = \left(y - \left(\beta_1 x + \beta_0\right)\right)^2$ then what is $\frac{\partial L}{\partial \beta_1}$?

$$L = \left(\left(y - \beta_1 x - \beta_0\right)^2\right)$$

where $\left(y - \beta_1 x - \beta_0\right)$ is f

$$\frac{\partial L}{\partial \beta_1} = \frac{\partial L}{\partial f}\frac{\partial f}{\partial \beta_0}$$

$$L = f^2$$

$$\frac{\partial L}{\partial f} = 2f$$

$$f = y - \beta_1 x - \beta_0$$

$$\frac{\partial f}{\partial \beta_1} = -x$$

$$\frac{\partial L}{\partial \beta_1} = \frac{\partial L}{\partial f}\frac{\partial f}{\partial \beta_1} = -2xf = -2x\left(y - \beta_1 x - \beta_0\right)$$

Optimization

In order to minimize the loss function, we end up with two equations, one being the partial derivative with respect to β_0 and the other with respect to β_1. Where the partial derivatives are zero, minimizes the loss function.

$$\nabla\left(L\right) = \left[\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1}\right] = 0$$

$$\frac{\partial L}{\partial \beta_0} = -2\left(y - \beta_1 x - \beta_0\right) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -2x\left(y - \beta_1 x - \beta_0\right) = 0$$

Putting it through the algebraic machine, we get the values of β_0 and β_1 that minimize the loss function.

$$\hat{\beta}_1 = \frac{\sum i\left(xi - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_i\left(x_i - \bar{x}\right)^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Summary: Estimate of the regression coefficients

We use MSE as our loss function,

$$L\left(\beta_0, \beta_1\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\left[y_i - \left(\beta_1 x_i + \beta_0\right)\right]\right)^2$$

We choose β_0 and β_1 to minimize the predictive errors made by our model. i.e. minimize our loss function.

Then the optimal values for $\hat{\beta}_0$ and $\hat{\beta}_1$ should be:

$$\hat{\beta}_0, \hat{\beta}_1 = \arg\min L\left(\beta_0, \beta_1\right)$$

β_0, β_1

Remember, we call this **fitting** or **training** the model.

Estimate of the regression coefficients: analytical solution

We calculate the partial derivatives and equate them to zero. Using algebra we find the values of β_0 and β_1.

Finding the exact solution only works in rare cases. Linear regression is one of these rare cases.

Take the gradient of the loss function and find the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ where the gradient is zero:

$$\nabla L = \left[\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1}\right] = 0$$

$$\hat{\beta}_1 = \frac{\sum i\,(xi - \bar{x})\,(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

where $\bar{y}$ and $\bar{x}$ are sample means.

The line which uses the β_0 and β_1

$$\hat{Y} = \hat{\beta}_1 X + \hat{\beta}_0$$

is called the regression line.

‹ Previous Next ›

Calculator