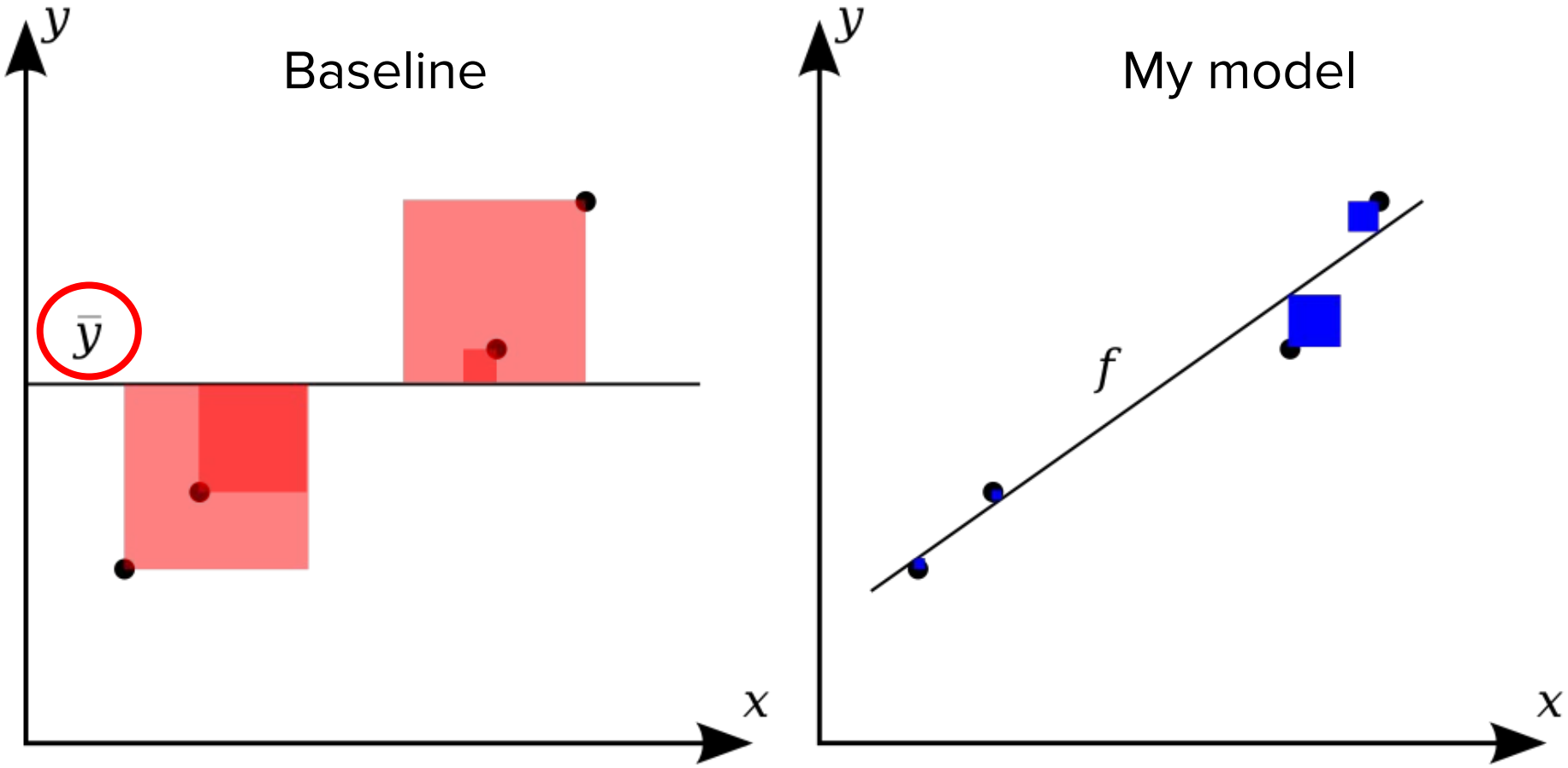
R2, MLR

Week 04 - Day 01

R2, R Squared, Coefficient of Determination

*“In statistics, the **coefficient of determination**, denoted R^2 or r^2 and pronounced "R squared", is the proportion of the variance in the dependent variable that is predictable from the independent variable(s).”*

Wikipedia


Baseline
y
$\bar{y}$
x
My model
y
f
x

What is blue_area / red_area?

blue_area == red_area

Meaning?

$$blue\_area < red\_area$$

Meaning?

blue_area > red_area

Meaning?

$$R2 = 1 - (blue\_area / red\_area)$$

Multiple Linear Regression

```
Salary =
+ b0
+ b1 * years_of_experience
+ b2 * ability_to_negotiate
- b3 * is_startup
+ b4 * is_finance
+ b5 * responsibilities
```

New assumptions:

independence of predictors

Multicollinearity

```
Salary =
+ b0
+ b1 * years_of_experience
+ b2 * months_of_experience
+ b3 * days_of_experience
```

Interpretation

```
Salary =
+ b0
+ b1 * years_of_experience
+ b2 * ability_to_negotiate
- b3 * is_startup
+ b4 * is_finance
+ b5 * responsibilities
```

What is b1?

Optimization Using Derivative

$$\text{Prediction} = b0 + (b1 * \text{input})$$

$$\text{Time} = 10 + 3 * \text{mrt\_stops}$$

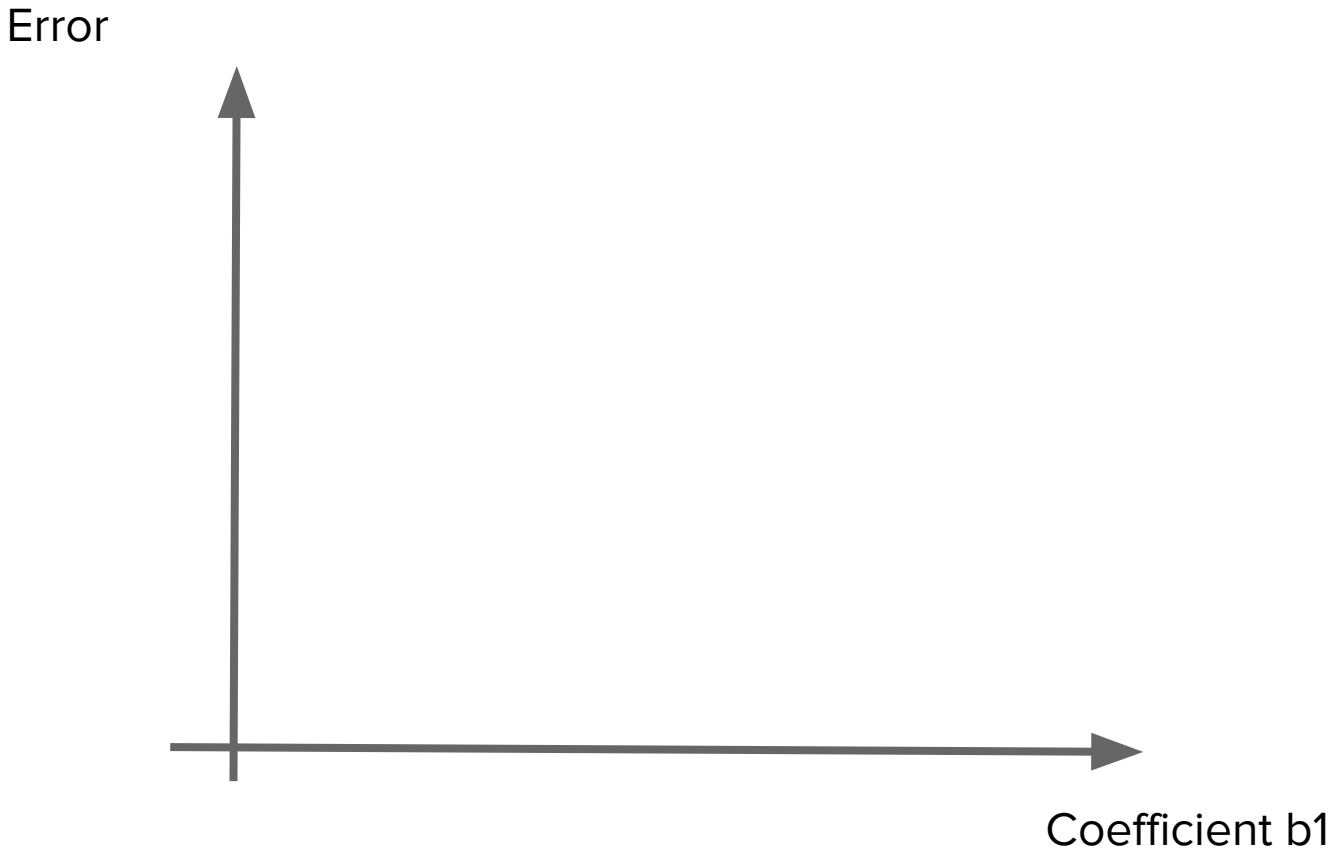
Optimization

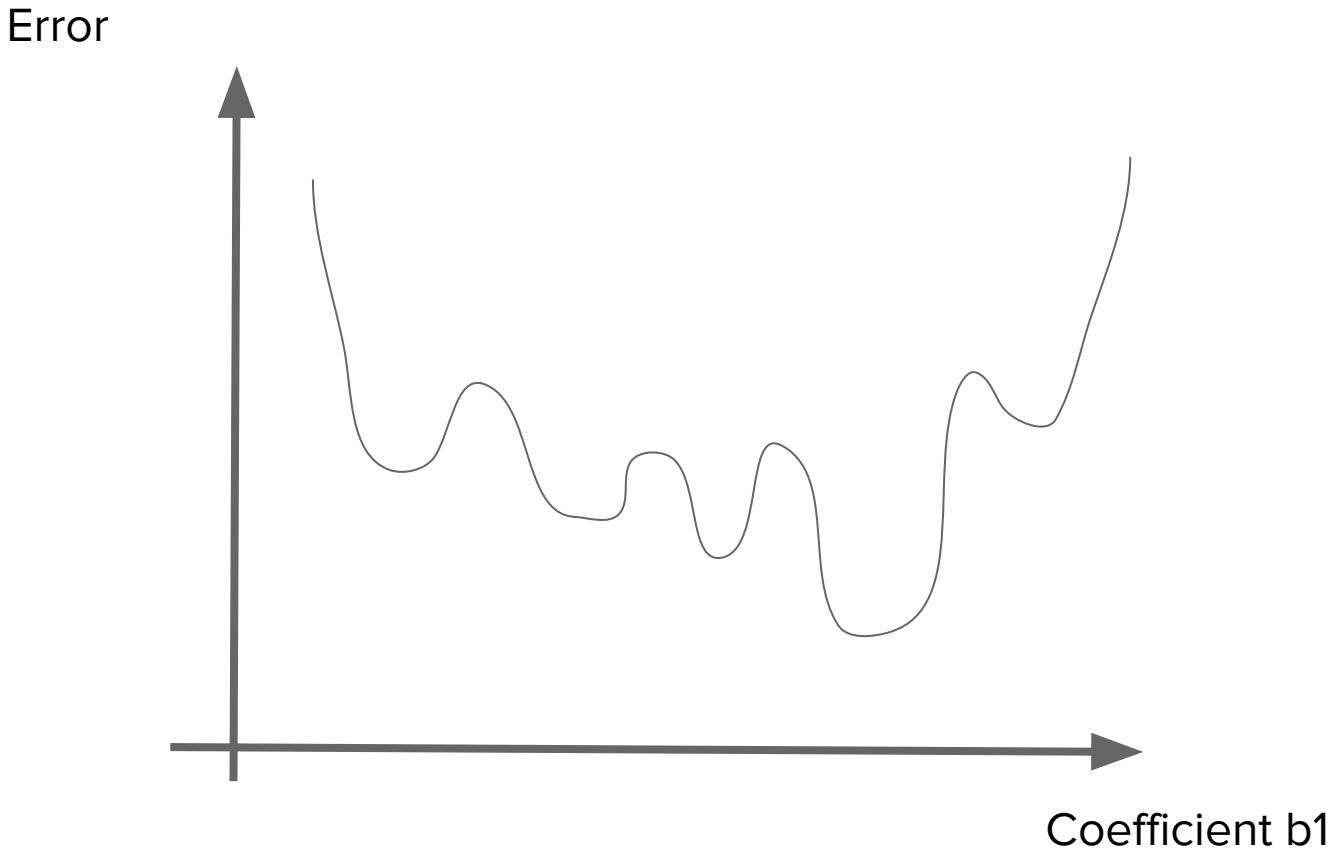
=

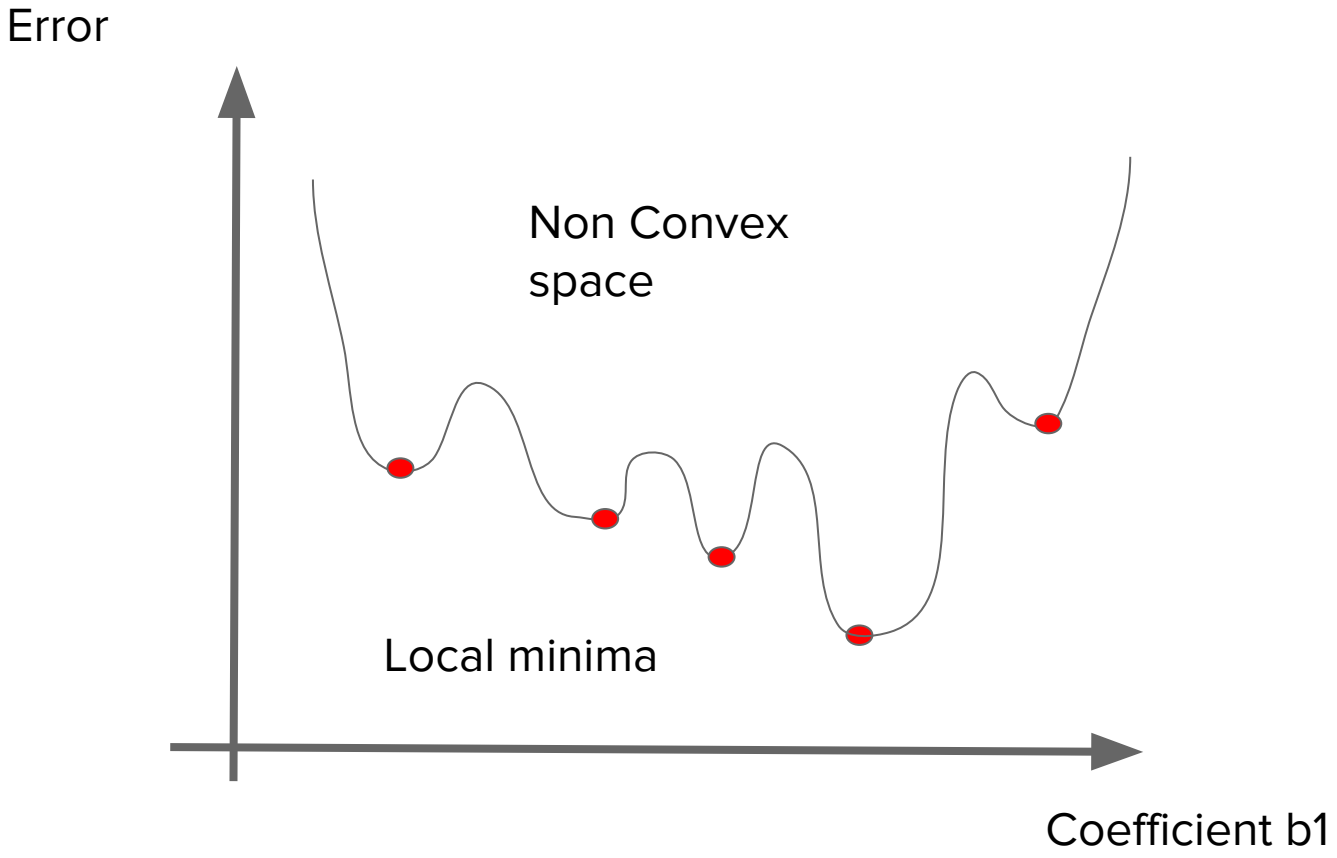
Finding b0,b1 that give us the smallest error

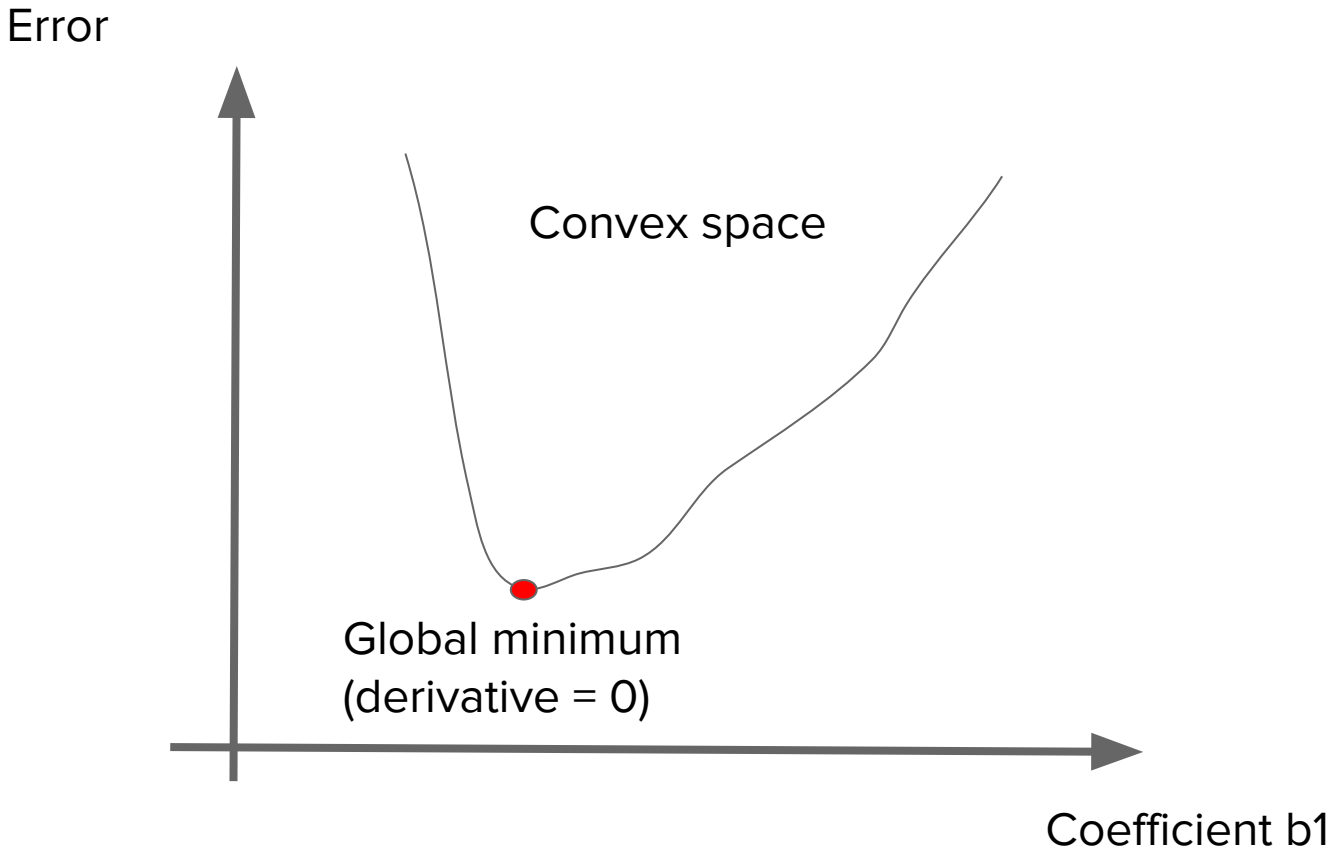
Derivative = minimum

NICE GIF ON WIKIPEDIA


Error
Coefficient b1


Error
Coefficient b1


Error
Non Convex
space
Local minima
Coefficient b1


Error
Convex space
Global minimum
(derivative = 0)
Coefficient b1

Quick note on matrices

Matrices on python = Numpy arrays

Solve the exercise

or

Review matrix multiplication, transpose, inverse, dot product, array-matrix multiplication, etc.