RMSLE: Cost Function

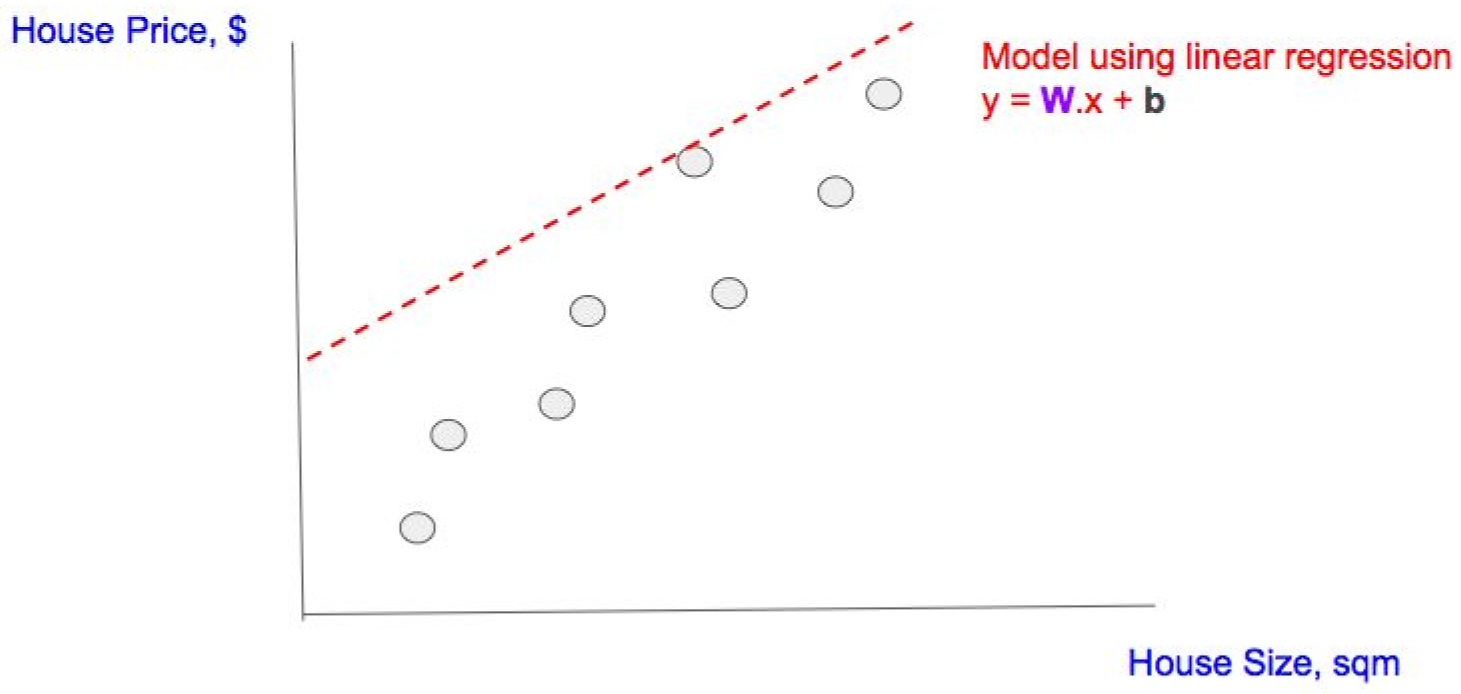
Khor Soon Hin, @neth_6, re:Culture
Katerina Malahova, Tokyo ML Gym

Machine Learning (ML) Lightning Tour

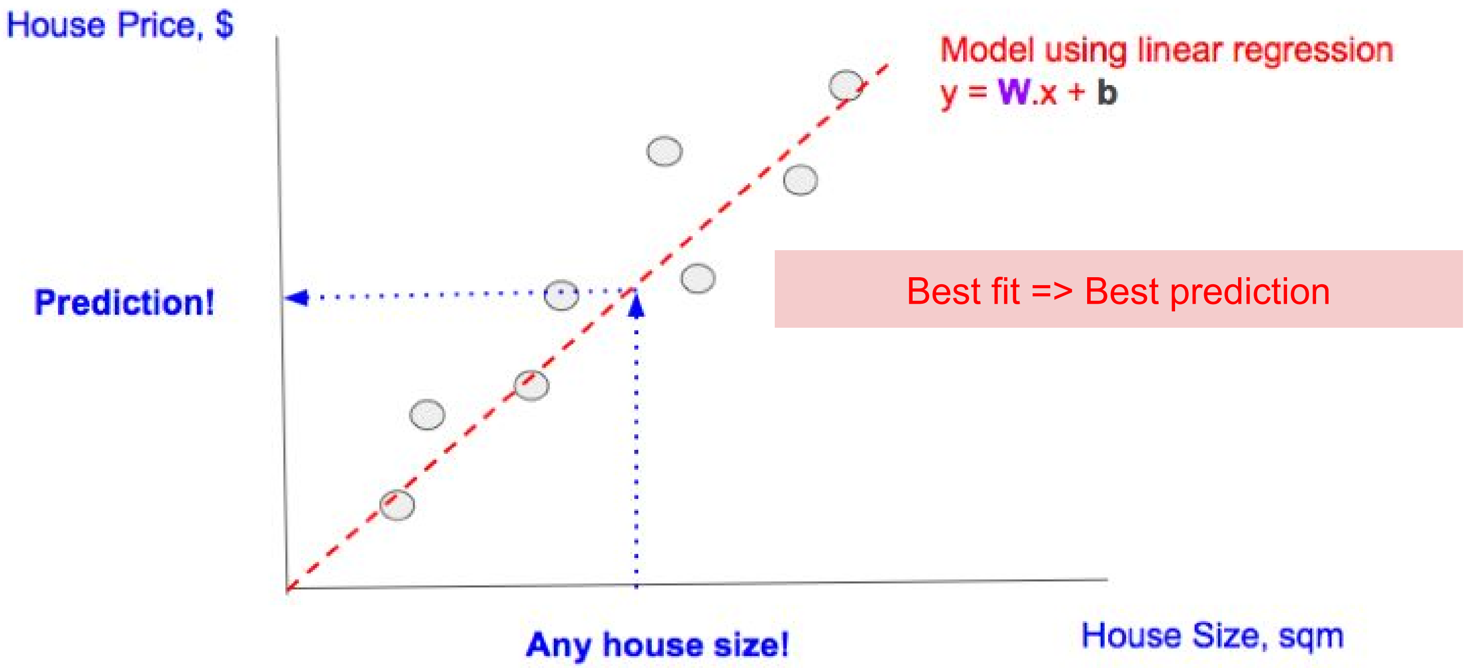
- Linear Regression in ML
- The role of Cost Function

Predicting using ML

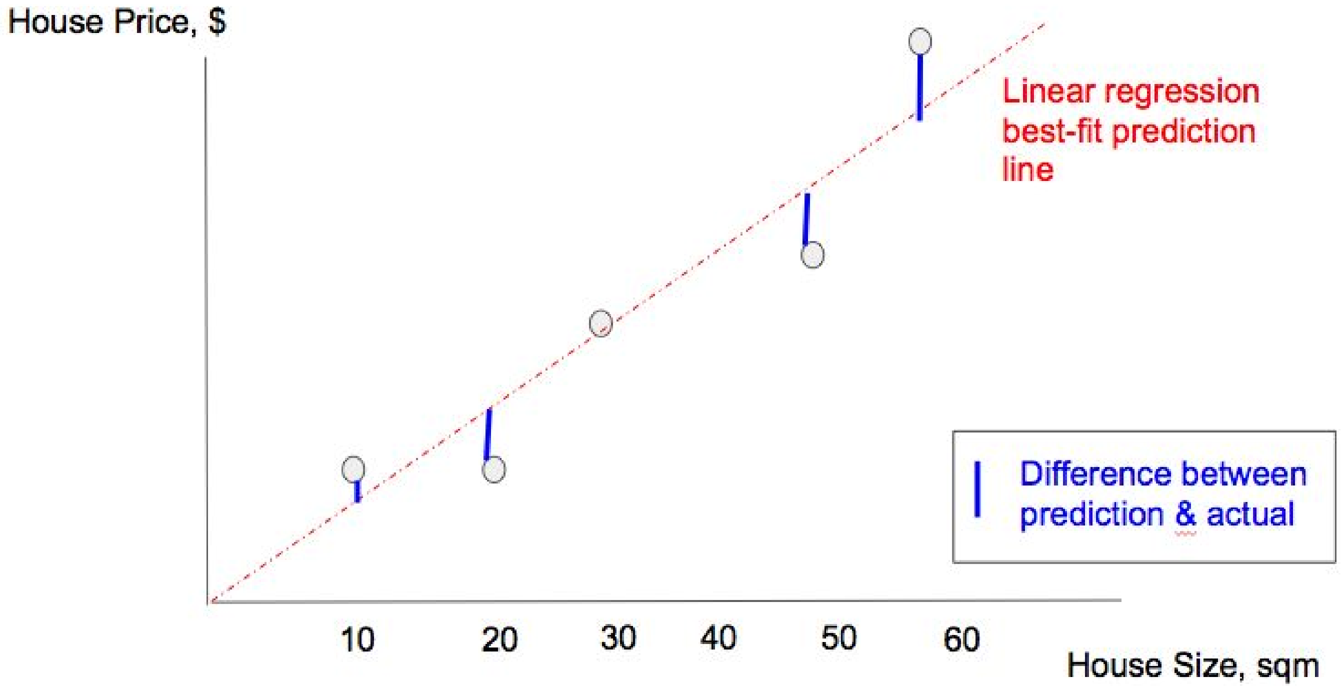
Linear Regression



Best Fit


House Price, $
Model using linear regression
y = W.x + b
Prediction!
Best fit => Best prediction
Any house size!
House Size, sqm

Best Fit Definition: Graphically



Cost Functions

Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\boxed{(y_i - \hat{y}_i)}^{\boxed{2}}}$$

Square to ensure always positive

Cost Functions

Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\boxed{(y_i - \hat{y}_i)}^2}$$

Root Mean Squared Log Error (RMSLE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\boxed{(\log(p_i + 1) - \log(a_i + 1))}^2}$$

prediction

actual

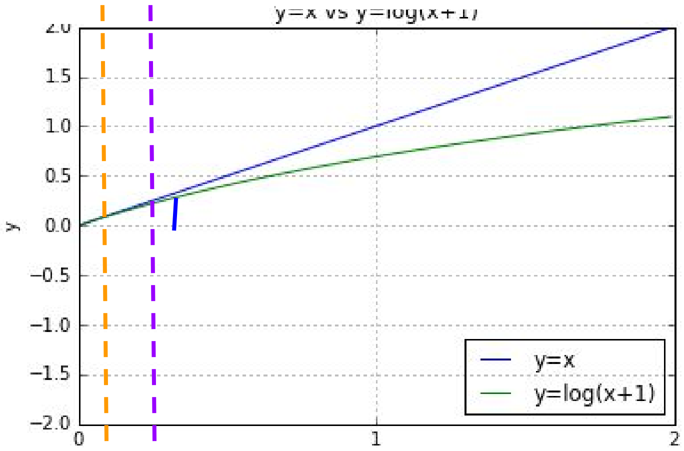
Cost Functions

Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boxed{y_i}-\boxed{\hat{y}_i})^2}$$

Root Mean Squared Log Error (RMSLE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(\boxed{p_i+1}) - \log(\boxed{a_i+1}))^2}$$



When predicted and actual is small:

For the same predicted & actual, RMSE & RMSLE is same (the blue vertical line)

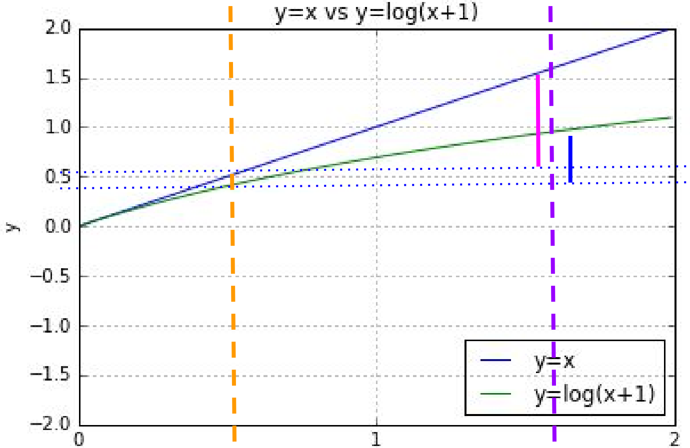
Cost Functions

Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Root Mean Squared Log Error (RMSLE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i+1) - \log(a_i+1))^2}$$



If either predicted or actual is big:

For the same predicted & actual, RMSE > RMSLE is same (the pink vs blue vertical line)

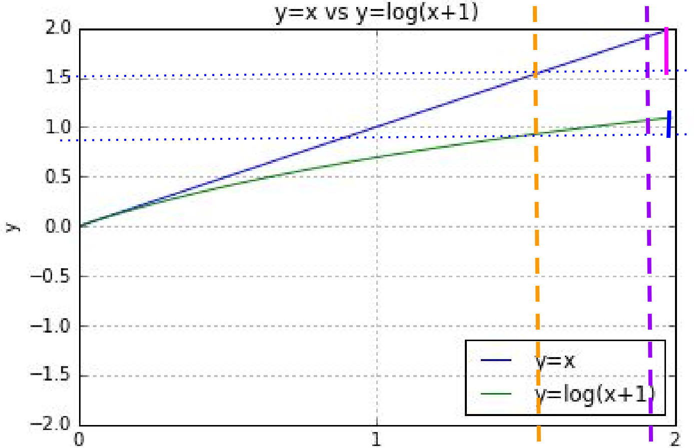
Cost Functions

Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Root Mean Squared Log Error (RMSLE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i + 1) - \log(a_i + 1))^2}$$



If both predicted and actual are big:

For the same predicted & actual, RMSE > RMSLE is same (the pink vs blue vertical line)

NOTE: RMSLE is almost negligible

Alternative: RMSLE Intuition

RMSLE: log(Pi + 1) - log(Ai +1) = log((Pi + 1)/(Ai +1))

Only the percentual differences matter!

For example for P = 1000 and A = 500 would give you the roughly same error as when P = 100000 and A = 50000

RMSLE is usually used when you don't want to penalize huge differences in the predicted and true values when both predicted and true values are huge numbers.

Credits to Katerina Malahova for sharing this

RMSLE Usage

- When prediction error for small prediction is undesirable
- Example:
    - Predict inventory required to fulfill all customers
        - When you are small, you have few customers so ensuring each one is satisfied is more important missing out a few customers when you are big
    - Kaggle: https://www.kaggle.com/wiki/RootMeanSquaredLogarithmicError

Cost Function

- Purpose:
  - Mathematical definition of best fit
  - ML’s goal is to minimize it
- Usage:
  - Used by ML to determine how to tweak model parameters to get best fit
    - Gradient descent