LEC 4: Logistic Regression

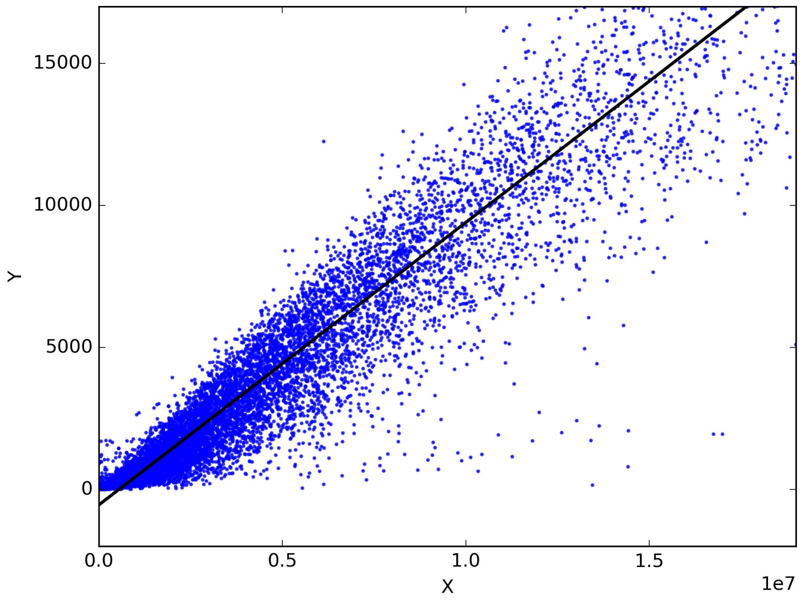
Feb 19, 2020

Our “ML” Algorithms up till now

Focus on Regression (Ordinary least squares, Ridge, and LASSO): Given k features X1, ..., Xk we predict the value of response variable Y

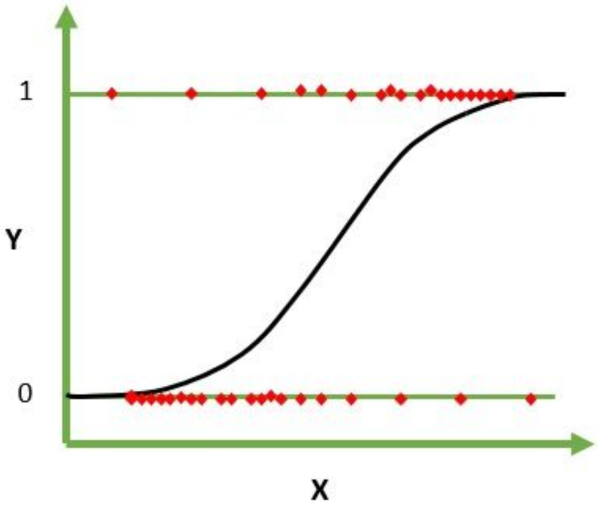
Y is continuous; it can take on any number of values

- Ex. predict value of house from # of windows, square footage



Introducing Logistic Regression

- What if Y can only take on two values: {0, 1}?
  - Ex. classify emails as spam (1) vs not spam (0)
  - Ex. classify an image as containing a car (1) vs not a car (0)
- This brings us to a new use of ML: classification
- Logistic Regression: given k features X1, ..., Xk predict probability of being 1

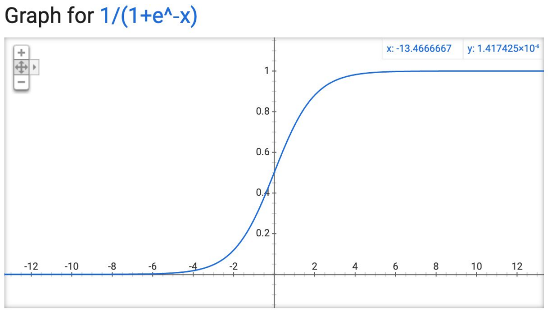


Basic idea of logistic regression

- Organize training data. Classify response variable as 0 or 1 (1 = email is spam, 0 = email is ham, not spam)
- Formulate our model using sigmoid function (what are we assuming here?):

$$P(Y_i = 1) = \frac{1}{1 + e^{-X_i^T \beta}} = \frac{1}{1 + e^{\beta_0 + \beta_1 X_{1i} + \ldots + \beta_1 X_{ki}}}$$

- A note on sigmoid function: $\sigma(x) = \dfrac{1}{1 + e^{-x}}$
- Find optimal values of $\beta_1, \ldots, \beta_k$
- Use model for predicting P(Y=1), classify



Why not just fit a OLS regression model?

- Non-constant variance of error terms (ui)
- Output value of ordinary least squares regression, Yi, isn’t very useful--it doesn’t mean anything
- On the other hand, logistic regression outputs the probability(observation i is = 1)
    - Much more interpretable