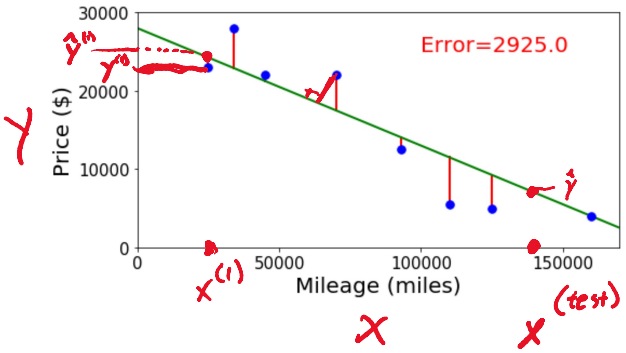
315 Lecture 2

- Data / examples / experience
  - Input
  - Output (labels)
- Model
- Parameters
  - Model complexity
- Hypothesis function
  - Prediction
- Error/loss, accuracy
- Objective function
- Global/local min/max
- Training, validation, test set
- Over (under) fitting
- Classification
- Regression
- Supervised (unsupervised) learning

Specific

- Linear model
- Mean squared error
- Gradient
- Gradient Descent
- Stochastic Gradient Descent
- Learning rate
- Batch
- Sigmoid
- ReLU

- Softmax
- Cross entropy
- Neural network



$y = mx + b$

$\hat{y}^{(test)} = m x^{(test)} + b$

$\hat{y} = h_\theta(x)$

$\hat{y}^{(i)} = m x^{(i)} + b$

$error^{(i)} = y^{(i)} - \hat{y}^{(i)}$

$\sum_{i=1}^{N} error^{(i)}$

$MSE = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - \hat{y}^{(i)})^2$

$loss(y, \hat{y}) = \frac{1}{N} \sum (y^{(i)} - \hat{y}^{(i)})^2$

Obj $J(\vec{\theta}) = loss(y, \hat{y})$

$= \frac{1}{N} \sum_i (y^{(i)} - h_\theta(x^{(i)}))^2$

$J(m, b) = \frac{1}{N} \sum_i (y^{(i)} - (m x^{(i)} + b))^2$