Session 3

Spring 1400 by Matin Zivdar

Table of Contents

Train/Dev/Test

layers

hidden units

Learning rates

Activation functions

...

Train/Dev/Test (cont.)

Data

Train Set

Dev Set
(aka Validate Set)

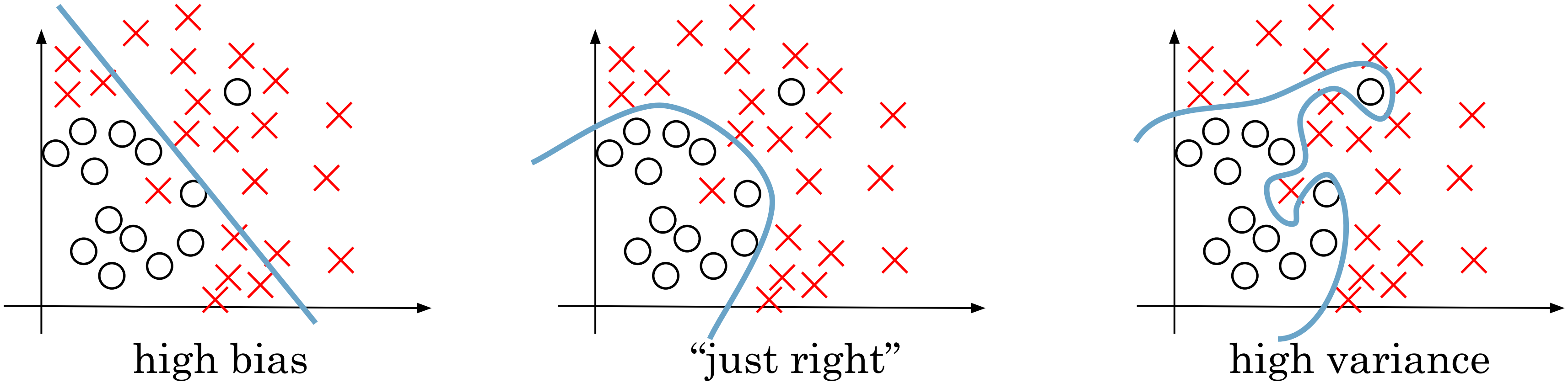
Train Set

Split Ratio?

50%-25%-25% / 60%-20%-20%

98%-1%-1% For big dataset in order of 1,000,000

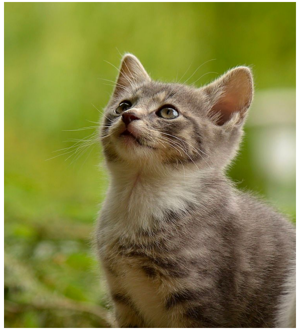
Bias and Variance

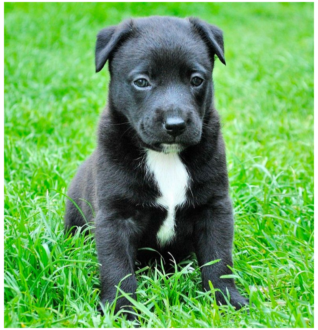


underfitting

overfitting

Bias and Variance (cont.)

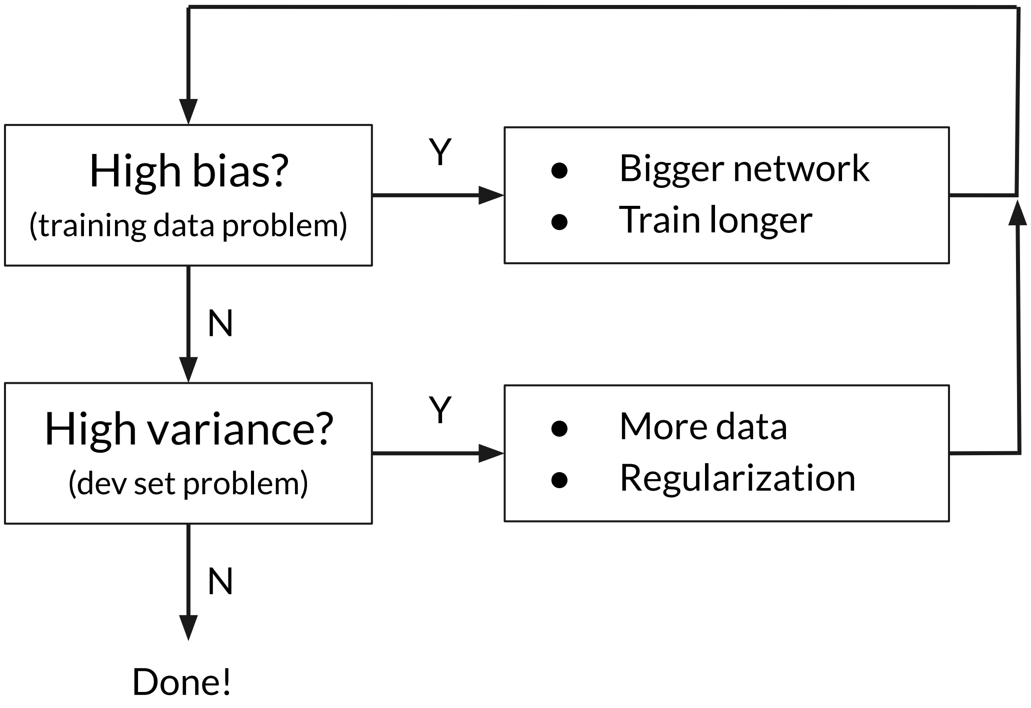




Cat classification

Dev set error:	1%	15%	15%	0.5%
Train set error:	11%	16%	30%	1%
	High Variance	High Bias	High Bias & High Variance	Low Bias & Low Variance

Basic recipe for machine learning



Regularization

- L2 Regularization
- Dropout Regularization

Logistic Regression

$$\min_{w,\, b} J(w,\ b)$$

$$w \in \mathbb{R}^{n_x},\ b \in \mathbb{R}$$

$$J(w,\ b) = \frac{1}{m}\sum_{i=1}^{m}\mathcal{L}\left(\hat{y}^{(i)}, y^{(i)}\right) + \frac{\lambda}{2m}\|w\|_2^2$$

$$\|w\|_2^2 = \sum_{j=1}^{n_x} w_j^2 = w^T.w$$

Neural Network

$$J\left(w^{[1]}, b^{[1]}, \dots, w^{[l]}, b^{[l]}\right) = \frac{1}{m}\sum_{i=1}^{m}\mathscr{L}\left(\hat{y}^{(i)}, y^{(i)}\right) + \frac{\lambda}{2m}\sum_{l=1}^{L}\| w^{[l]} \|_F^2$$

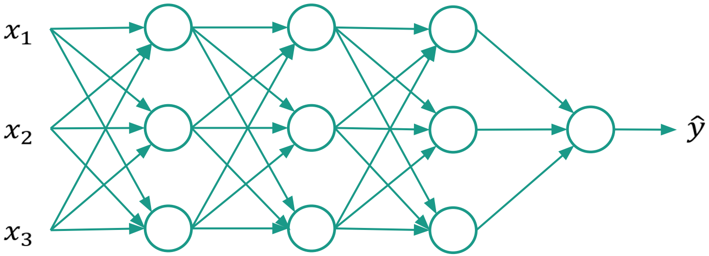
$$\| w^{[l]} \|_F^2 = \sum_{i=1}^{n[l-1]}\sum_{j=1}^{n[l]}\left(w_{ij}^{[l]}\right)^2$$

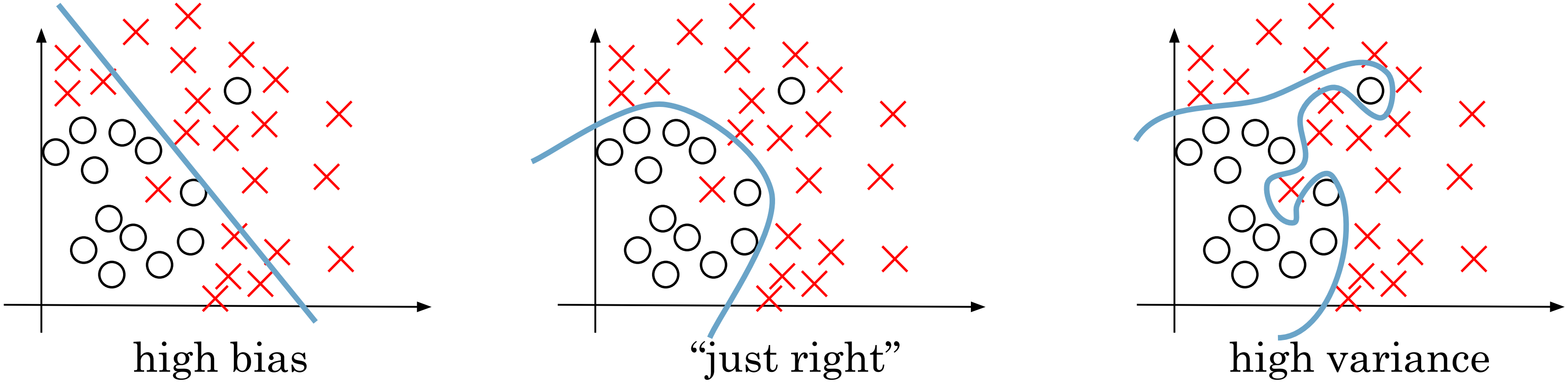
$$dw^{[l]} = (\text{from backprop}) + \frac{\lambda}{m} w^{[l]}$$

$$w^{[l]} = w^{[l]} - \alpha \, dw^{[l]}$$

Why does it work?

$$J = \frac{1}{m}\sum_{i=1}^{m}\mathscr{L}\left(\hat{y}^{(i)}, y^{(i)}\right) + \frac{\lambda}{2m}\sum_{l=1}^{L}\left\| w^{[l]} \right\|_F^2$$





Dropout Regularization

No dropout at Test time!