Deep Learning: Theory and Practice

Linear and Logistic Models for Classification

14-02-2019

deeplearning.cce2019@gmail.com

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Logistic Regression

- 2- class logistic regression $$p(\mathcal{C}_1|\phi) = y(\phi) = \sigma\left(\mathbf{w}^{\mathrm{T}}\phi\right)$$

- Maximum likelihood solution $$\nabla E(\mathbf{w}) = \sum_{n=1}^{N}(y_n - t_n)\phi_n$$

- K-class logistic regression $$p(\mathcal{C}_k|\phi) = y_k(\phi) = \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

$$a_k = \mathbf{w}_k^{\mathrm{T}}\phi.$$

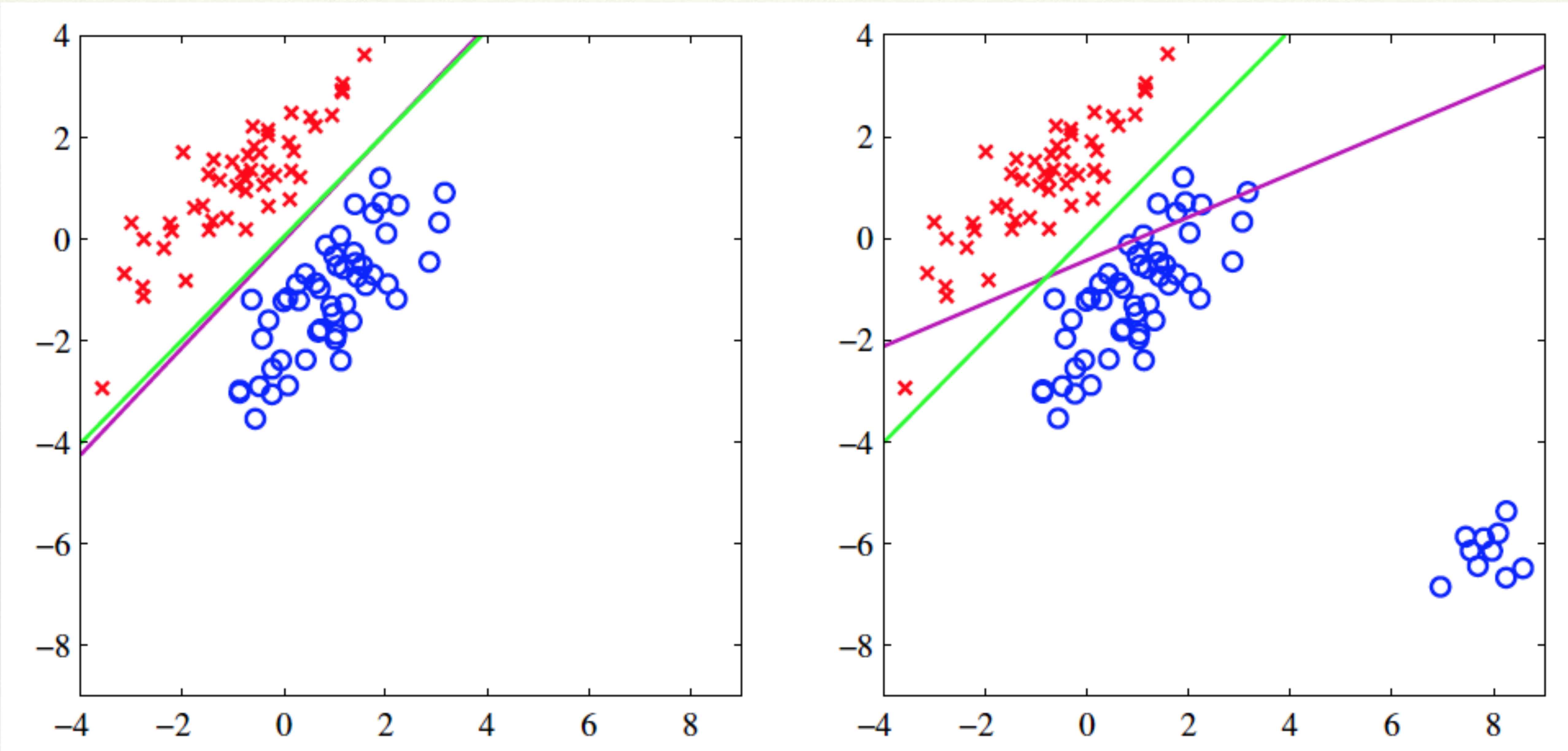
- Maximum likelihood solution

$$\nabla_{\mathbf{w}_j} E(\mathbf{w}_1, \ldots, \mathbf{w}_K) = \sum_{n=1}^{N}(y_{nj} - t_{nj})\,\phi_n$$

Bishop - PRML book (Chap 3)

Established 1911

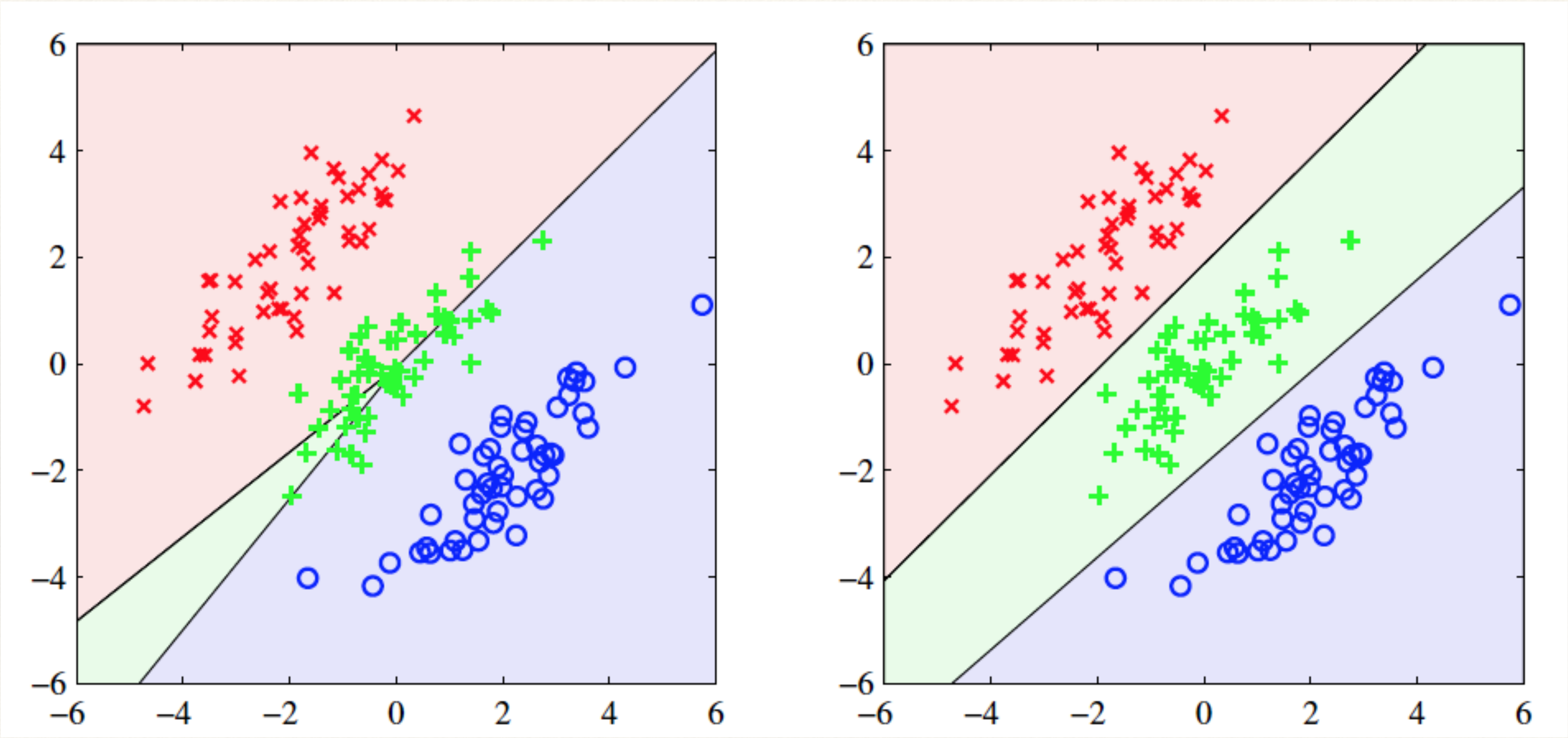
Least Squares versus Logistic Regression



Bishop - PRML book (Chap 4)

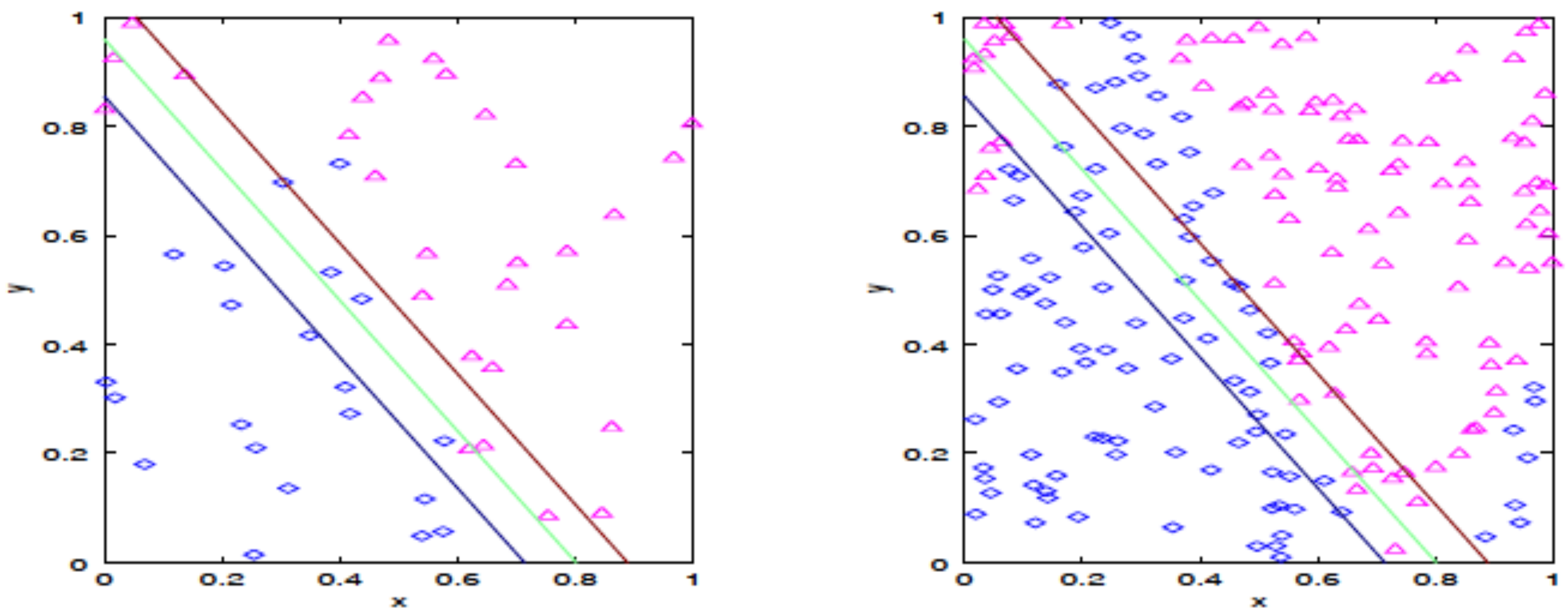
Established 1911

Least Squares versus Logistic Regression



Bishop - PRML book (Chap 4)

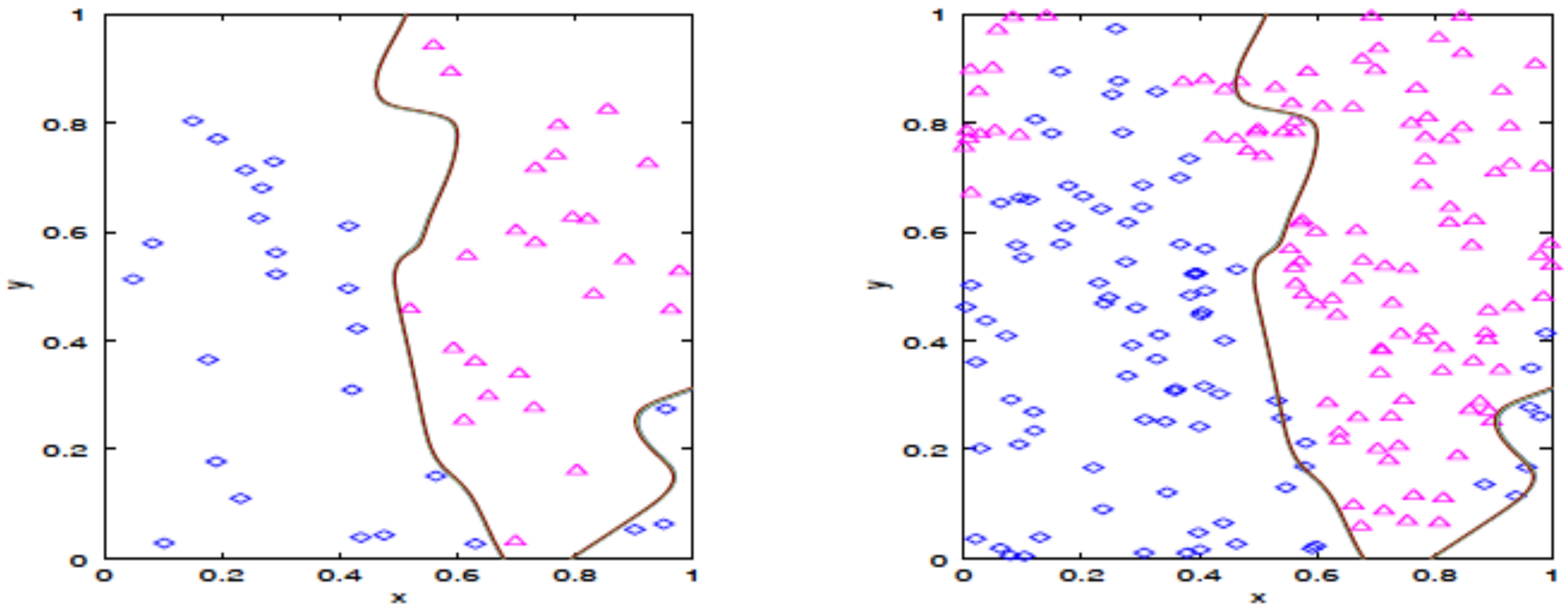
Underfit



- The model is not able to capture the variability in the data (Linear Model)
- Both the training and testing error are high (15%,20%)
- Try to learn a more complex model – more features, more hidden neurons, decrease regularization
- More data would not help

(Courtesy - Dr D. Vijayasenan, NITK)

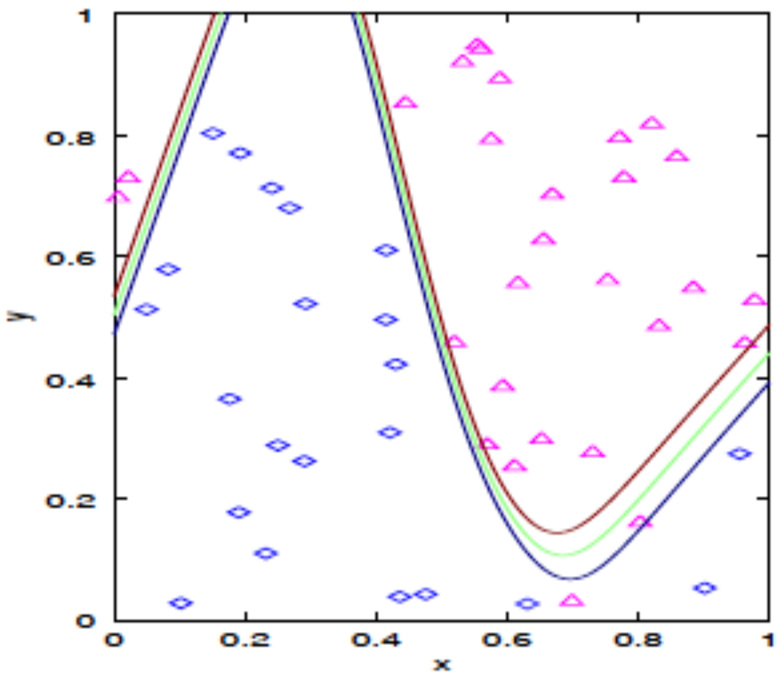
Overfit

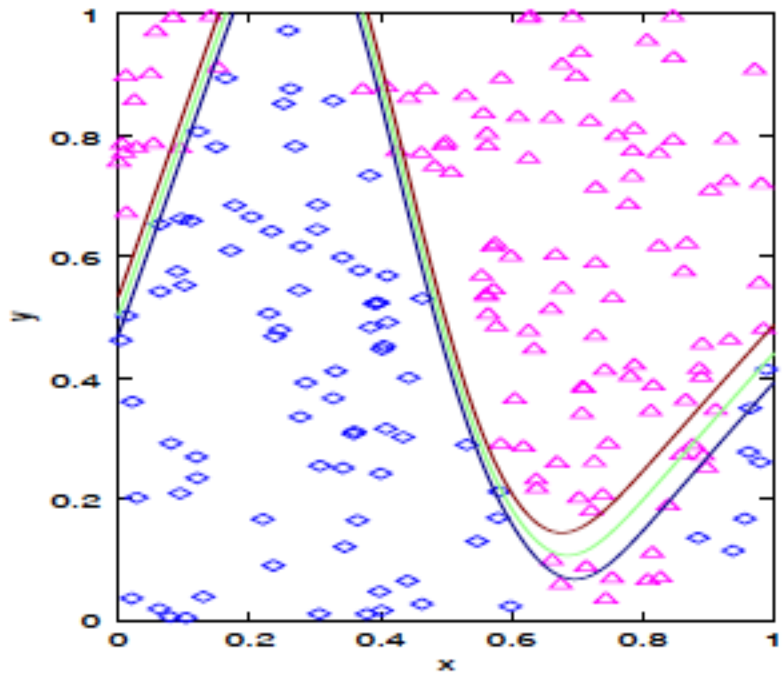


- The model is capturing data as well as accidental variations (100 hidden neurons)
- Training error is too low and testing error is too high (0%, and 16%)
- Try to learn a simpler model – less features, less hidden neurons, increase regularization
- More data would help

(Courtesy - Dr D. Vijayasenan, NITK)

Compromise





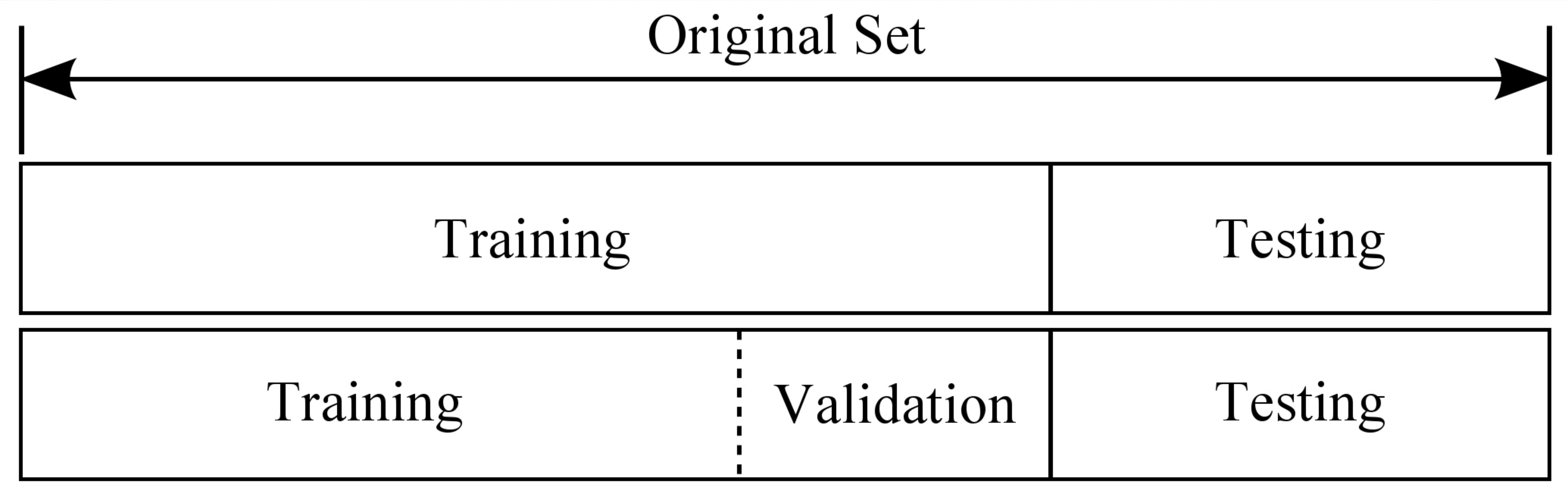
- Reasonable training and test errors – (4%, 8%)
- Appropriate model – capturing only the global characteristics not details

(Courtesy - Dr D. Vijayasenan, NITK)

Summary so far …

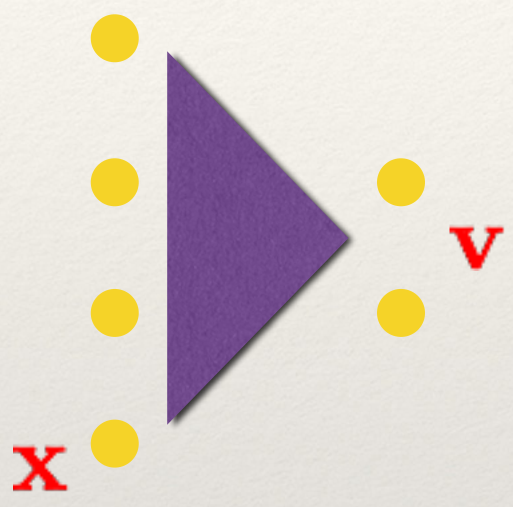
- Maximum Likelihood
- Linear Least Squares Classifiers
- Logistic Regression
  - Application of ML to Logistic Regression
  - Gradient Descent
  - Coding Logistic regression

Training, Validation and Test Set



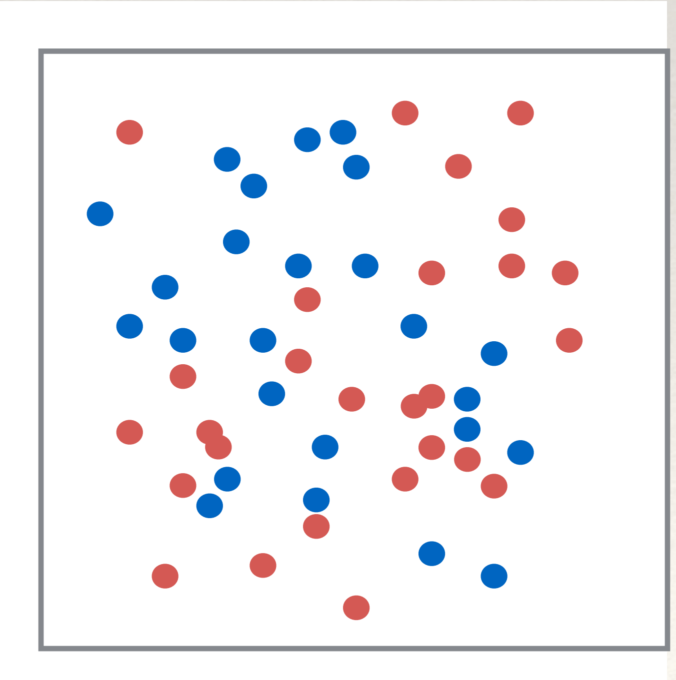
Perceptron Algorithm

Perceptron Model [McCulloch, 1943, Rosenblatt, 1957]



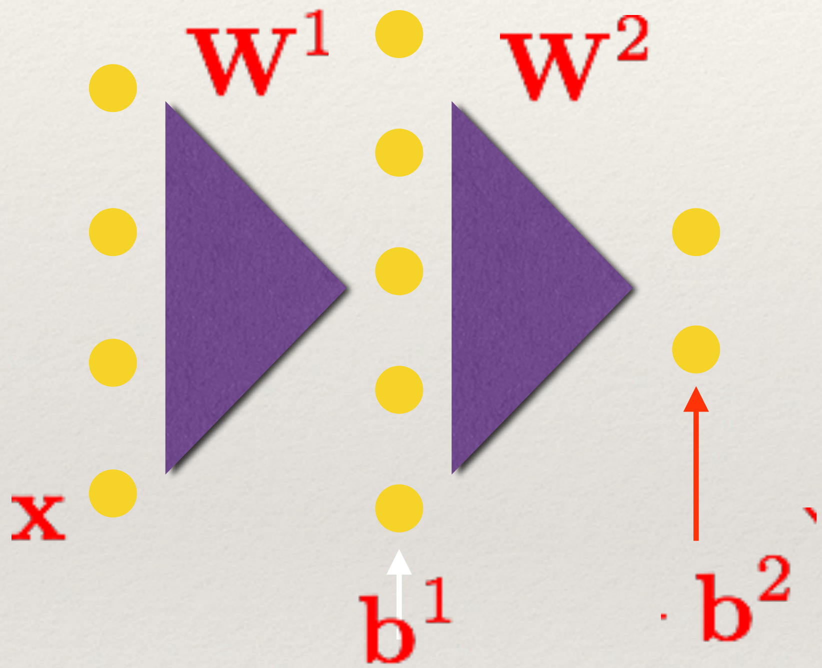
Targets are binary classes [-1,1]

What if the data is not linearly separable



Multi-layer Perceptron

Multi-layer Perceptron [Hopfield, 1982]



$$\mathbf{v}^2 = \psi\left(\mathbf{W}^2\phi(\mathbf{W}^1\mathbf{x} + \mathbf{b}^1) + \mathbf{b}^2\right)$$

ϕ non-linear function (*tanh*,*sigmoid*)

ψ thresholding function