Quick Recap for Linear models

Linear Regression: $y = \boldsymbol{w}^T\overline{\boldsymbol{x}}$

Linear classification

Mixture of Gaussians

Prior: $\Pr(Y) = \pi$ where π is a probability

Likelihood: $\Pr(\boldsymbol{X}|Y) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$)

Posterior: $\Pr(Y|\boldsymbol{X}) = kPr(Y)\Pr(\boldsymbol{X}|Y)$

Logistic Regression

$\Pr(Y|\boldsymbol{X}) = \sigma(\boldsymbol{w}^T\overline{\boldsymbol{x}})$ where σ is the logistic function for binary classification

$\Pr(Y_k|\boldsymbol{X}) = \frac{\exp\left(\boldsymbol{w}_k^T\overline{\boldsymbol{x}}\right)}{\sum_j \exp\left(\boldsymbol{w}_j^T\overline{\boldsymbol{x}}\right)}$ softmax function for multiclass classification

Thresholded perceptron

$y = sign(\boldsymbol{w}^T\overline{\boldsymbol{x}})$ where sign returns 1 when $\boldsymbol{w}^T\boldsymbol{x} \geq 0$ and 0 otherwise

Logistic perceptron

$\Pr(y) = \sigma(\boldsymbol{w}^T\overline{\boldsymbol{x}})$

Quick recap for non-linear models

Main idea use basis functions $\phi(\boldsymbol{x})$ to transform the input space in a non-linear fashion

Example: to span a polynomial of degree k, use $\phi_1(x) = x, \phi_2(x) = x^2, \ldots, \phi_k(x) = x^k$

$$x = 2 \quad \rightarrow \quad \phi(x) = \begin{pmatrix} x \\ x^2 \\ \cdots \\ x^k \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ \cdots \\ 2^k \end{pmatrix}$$

Non-linear regression: $y = \boldsymbol{w}^T \phi(\overline{\boldsymbol{x}})$

Multilayer neural network

Nonlinear classification

Logistic Regression

$\Pr(Y|\boldsymbol{X}) = \sigma(\boldsymbol{w}^T \phi(\overline{\boldsymbol{x}}))$ where σ is the logistic function for binary classification

$\Pr(Y_k|\boldsymbol{X}) = \frac{\exp(\boldsymbol{w}_k^T \phi(\overline{\boldsymbol{x}}))}{\sum_j \exp\left(\boldsymbol{w}_j^T \phi(\overline{\boldsymbol{x}})\right)}$ softmax function for multiclass classification

Thresholded perceptron

$y = sign(\boldsymbol{w}^T \phi(\overline{\boldsymbol{x}}))$ where sign returns 1 when $\boldsymbol{w}^T \phi(\overline{\boldsymbol{x}}) \geq 0$ and 0 otherwise

Logistic perceptron

$\Pr(y) = \sigma(\boldsymbol{w}^T \phi(\overline{\boldsymbol{x}}))$

Multi-layer neural networks