Machine Learning

Week 3: Simple Classifiers (Cont'd)

Mahesan Niranjan

School of Electronics and Computer Science
University of Southampton

Autumn Semester 2017/18

Outline

- Posterior probabilities for simple Gaussian cases
- Fisher Linear Discriminant
- Nearest Neighbour Classifier
- Classifier performance

Posterior probabilities for simple Gaussian cases

Two class problem

- Bayes classifier:

$$P\left[\omega_1|\boldsymbol{x}\right] = \frac{p\left(\boldsymbol{x}|\omega_1\right) P\left[\omega_1\right]}{p\left(\boldsymbol{x}|\omega_1\right) P\left[\omega_1\right] + p\left(\boldsymbol{x}|\omega_2\right) P\left[\omega_2\right]}$$

- Restrictive assumptions:
  - Gaussian $p\left(\boldsymbol{x}|\omega_j\right) = \mathcal{N}\left(\boldsymbol{m}_j, \boldsymbol{C}_j\right)$
  - Equal covariance matrices: $\boldsymbol{C}_1 = \boldsymbol{C}_2 = C$
- Substitute, divide through by numerator term and cancel common terms to get

$$P\left[\omega_1|\boldsymbol{x}\right] = \frac{1}{1 + \exp\left\{-(\boldsymbol{w}^t \boldsymbol{x} + w_0)\right\}}$$

- The functional form $1/(1 + \exp(-\alpha))$ is known as sigmoid / logistic (See lab class for W3)

Fisher Linear Discriminant

- Classification problem (say two classes)
- Disirable properties of a direction to project
  - Means of projected data should be far apart
  - Variance of projections of each class should be small

Fisher Linear Discriminant

- In the $p-$ dimensional ($\mathcal{R}^p$) input space, find a direction on which projected data is maximally separable:
  - Projected means should be far apart
  - Projected scatter of each class should be small
- Projection of $\boldsymbol{x}_n$ onto direction $\boldsymbol{w}$ is $\boldsymbol{w}^t\boldsymbol{x}_n$;
  - Projected mean for class j will be at $\boldsymbol{w}^t\boldsymbol{m}_j$
  - Variance of projections is $\boldsymbol{w}^t\boldsymbol{C}_j\boldsymbol{w}$, where $\boldsymbol{C}_j$ is the covariance matrix of data in class j.
- Fisher Ratio:

$$J_F = \frac{(\boldsymbol{w}^t\boldsymbol{m}_1 - \boldsymbol{w}^t\boldsymbol{m}_2)^2}{\boldsymbol{w}^t\boldsymbol{C}_1\boldsymbol{w} + \boldsymbol{w}^t\boldsymbol{C}_2\boldsymbol{w}}$$

- We can write the numerator as $\boldsymbol{w}^t\boldsymbol{C}_B\boldsymbol{w}$, where $\boldsymbol{C}_B = (\boldsymbol{m}_1 - \boldsymbol{m}_2)(\boldsymbol{m}_1 - \boldsymbol{m}_2)^t$, the between-class scatter matrix.
- $\boldsymbol{C}_W = \boldsymbol{C}_1 + \boldsymbol{C}_2$, the within class scatter matrix.

Fisher Linear Discriminant (cont'd)

- Fisher criterion to maximize

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^t\,\boldsymbol{C}_B\,\boldsymbol{w}}{\boldsymbol{w}^t\,\boldsymbol{C}_W\,\boldsymbol{w}}$$

- Set gradient to zero

$$\nabla_{\boldsymbol{w}} = \frac{2\boldsymbol{C}_B\boldsymbol{w} \times \left(\boldsymbol{w}^t\,\boldsymbol{C}_W\,\boldsymbol{w}\right) - 2\boldsymbol{C}_W\boldsymbol{w} \times \left(\boldsymbol{w}^t\,\boldsymbol{C}_B\,\boldsymbol{w}\right)}{(\boldsymbol{w}^t\,\boldsymbol{C}_W\,\boldsymbol{w})^2}$$

- Equate this to zero and observe
  - $\boldsymbol{w}^t\boldsymbol{C}_W\boldsymbol{w}$ and $\boldsymbol{w}^t\boldsymbol{C}_B\boldsymbol{w}$ are scalars
  - $\boldsymbol{C}_B\boldsymbol{w}$ points in the same direction as $\boldsymbol{m}_1 - \boldsymbol{m}_2$
  - We are only interested in the direction of $\boldsymbol{w}$

$$\boldsymbol{w}_F = \alpha\boldsymbol{C}_W^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_2)$$