

Machine Learning

Week 3: Simple Classifiers (Cont'd)

Maheesan Niranjana

School of Electronics and Computer Science
University of Southampton

Autumn Semester 2016/17

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:

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

- Restrictive assumptions:

- Gaussian $p(\mathbf{x}|\omega_j) = \mathcal{N}(\mathbf{m}_j, \mathbf{C}_j)$
- Equal covariance matrices: $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{C}$

- Substitute, divide through by numerator term and cancel common terms to get

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

- 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)
- Desirable 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 \mathbf{x}_n onto direction \mathbf{w} is $\mathbf{w}^t \mathbf{x}_n$;
 - Projected mean for class j will be at $\mathbf{w}^t \mathbf{m}_j$
 - Variance of projections is $\mathbf{w}^t \mathbf{C}_j \mathbf{w}$, where \mathbf{C}_j is the covariance matrix of data in class j .

- Fisher Ratio:

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

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

Fisher Linear Discriminant (cont'd)

- Fisher criterion to maximize

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

- Set gradient to zero

$$\nabla_{\mathbf{w}} = \frac{2\mathbf{C}_B \mathbf{w} \times (\mathbf{w}^t \mathbf{C}_W \mathbf{w}) - 2\mathbf{C}_W \mathbf{w} \times (\mathbf{w}^t \mathbf{C}_B \mathbf{w})}{(\mathbf{w}^t \mathbf{C}_W \mathbf{w})^2}$$

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

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