ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG
INSTITUT FÜR INFORMATIK
Lehrstuhl für Mustererkennung und Bildverarbeitung
Prof. Dr.-Ing. Hans Burkhardt

Georges-Köhler-Allee
Geb. 052, Zi 01-029
D-79110 Freiburg
Tel. 0761 - 203 - 8260

Übungen zur Vorlesung
Grundlagen der Bilderzeugung und Bildanalyse (Mustererkennung)
WS 05/06

Musterlösung 13

Aufgabe 13.1: Regularisierte Lineare Regression

a) Unser Minimierungproblem lässt sich wie folgt formulieren:

$$J(\mathbf{w}) = E\left\{\left\|\mathbf{w}^T\mathbf{x} - \mathbf{y}\right\|^2\right\} + \alpha E\left\{\|\mathbf{w}\|^2\right\} \tag{1}$$

Bis auf den zweiten Term ist diese Formulierung gleich der vom Polynomklassifikator. Mit dem Variationsansatz (s. Kap. 9, Folie 14) ergibt sich:

$$J(\mathbf{w}^*+\delta\mathbf{w}) = 1 - 2Spur\left[(\mathbf{w}^*+\delta\mathbf{w})^T E\left\{\mathbf{x}\mathbf{y}^T\right\}\right] + Spur\left[(\mathbf{w}^*+\delta\mathbf{w})^T E\left\{\mathbf{x}\mathbf{x}^T\right\}(\mathbf{w}^*+\delta\mathbf{w})\right]$$
$$+\alpha(\mathbf{w}^*+\delta\mathbf{w})^T(\mathbf{w}^*+\delta\mathbf{w})$$

Wir betrachten nur den letzten Term. Dann gilt:

$$\alpha(\mathbf{w}^*+\delta\mathbf{w})^T(\mathbf{w}^*+\delta\mathbf{w}) = Spur\left[\alpha(\mathbf{w}^*+\delta\mathbf{w})(\mathbf{w}^*+\delta\mathbf{w})^T\right]$$
$$= Spur\left[\alpha\delta^2\mathbf{w}\mathbf{w}^T\right] - Spur\left[\alpha\delta\mathbf{w}^*\mathbf{w}^T\right] + Spur\left[\alpha\mathbf{w}^*\mathbf{w}^{*T}\right]$$

Und somit erhalten wir für $J(\mathbf{w}^*+\delta\mathbf{w})$:

$$J(\mathbf{w}^*+\delta\mathbf{w}) = 1 - 2Spur\left[\mathbf{w}^{*T}E\left\{\mathbf{x}\mathbf{y}^T\right\}\right] - 2Spur\left[\delta\mathbf{w}^T E\left\{\mathbf{x}\mathbf{y}^T\right\}\right]$$
$$+Spur\left[\mathbf{w}^{*T}E\left\{\mathbf{x}\mathbf{x}^T\right\}\mathbf{w}^*\right] + 2Spur\left[\delta\mathbf{w}^T E\left\{\mathbf{x}\mathbf{x}^T\right\}\mathbf{w}^*\right]$$
$$+Spur\left[\delta\mathbf{w}^T E\left\{\mathbf{x}\mathbf{x}^T\right\}\delta\mathbf{w}\right] + Spur\left[\alpha\delta^2\mathbf{w}\mathbf{w}^T\right] - 2Spur\left[\alpha\delta\mathbf{w}^*\mathbf{w}^T\right] + Spur\left[\alpha\mathbf{w}^*\mathbf{w}^{*T}\right]$$
$$= J(\mathbf{w}^*) + \underbrace{Spur\left[\delta\mathbf{w}^T E\left\{\mathbf{x}\mathbf{x}^T\right\}\delta\mathbf{w}\right]}_{pos. definit}$$
$$+\underbrace{Spur\left[\alpha\delta^2\mathbf{w}\mathbf{w}^T\right]}_{pos. definit} - 2Spur\left[\delta\mathbf{w}^T(E\left\{\mathbf{x}\mathbf{y}^T\right\} - E\left\{\mathbf{x}\mathbf{x}^T\right\}\mathbf{w}^* + \alpha\mathbf{I}_m\mathbf{w}^*\right]$$

Laut Variationsansatz muss für alle $\delta\mathbf{w} \neq 0$ gelten:

$$J(\mathbf{w}^*+\delta\mathbf{w}) > J(\mathbf{w}^*)$$

Der 2. und 3. Term sind positiv definit. Daher muss gelten:

$$E\left\{\mathbf{x}\mathbf{y}^T\right\} - E\left\{\mathbf{x}\mathbf{x}^T\right\}\mathbf{w}^* + \alpha\mathbf{I}_m\mathbf{w}^* = 0$$
$$\Leftrightarrow R_{xy} - R_{xx}\mathbf{w}^* + \alpha\mathbf{I}_m\mathbf{w}^* = 0$$
$$\Leftrightarrow R_{xy} = (R_{xx} + \alpha\mathbf{I}_m)\mathbf{w}^*$$
$$\Leftrightarrow \mathbf{w}^* = (R_{xx} + \alpha\mathbf{I}_m)^{-1}R_{xy}$$

b)Wir zerlegen ein beliebiges $\mathbf{w} \in I\!R^m$ in einen Anteil $\mathbf{w}_\parallel = \sum_{i=1}^n \beta_i \mathbf{x}_i$ mit $\beta_i \in I\!R$ und in dessen orthogonales Kompliment $\mathbf{w}_\perp$. Dann gilt:

$$\mathbf{w} = \mathbf{w}_\parallel + \mathbf{w}_\perp \tag{2}$$

$$\mathbf{w}_\parallel^T \mathbf{w}_\perp = 0 \tag{3}$$

Wir können nun $\mathbf{w}$ in (1) einsetzen und erhalten:

$$J(\mathbf{w}_\parallel + \mathbf{w}_\perp) = E\left\{\left\|(\mathbf{w}_\parallel + \mathbf{w}_\perp)^T \mathbf{x} - \mathbf{y}\right\|^2\right\} + \alpha E\left\{\left\|\mathbf{w}_\parallel + \mathbf{w}_\perp\right\|^2\right\}$$

Da $\mathbf{w}_\perp \mathbf{x}_i = 0$ ist, erhalten wir:

$$J(\mathbf{w}_\parallel + \mathbf{w}_\perp) = E\left\{\left\|\mathbf{w}_\parallel^T \mathbf{x} - \mathbf{y}\right\|^2\right\} + \alpha E\left\{\left\|\mathbf{w}_\parallel + \mathbf{w}_\perp\right\|^2\right\}$$

Dieser Ausdruck wird minmal für $\mathbf{w}_\perp = 0$ und somit:

$$\mathbf{w}^* = \mathbf{w}_\parallel = \sum_{i=1}^n \beta_i \mathbf{x}_i$$

c)Wir setzen in (1) $\mathbf{w} = \sum_i^n \beta_i \mathbf{x}_i$. Dann ist:

$$J(\sum_i^n \beta_i \mathbf{x}_i) = \sum_j^n \left|(\sum_i^n \beta_i \mathbf{x}_i)^T - \mathbf{y}_j\right|^2 + \alpha(\sum_i^n \beta_i \mathbf{x}_i)^T(\sum_i^n \beta_i \mathbf{x}_i)$$

$$= \sum_j^n \left|(\sum_i^n \beta_i \mathbf{x}_i)^T \mathbf{x}_j - \mathbf{y}_j\right|^2 + \alpha \sum_j^n \left|\beta_j(\sum_i^n \beta_i \mathbf{x}_i)\mathbf{x}_j\right|^2$$

$$= \left\|(\sum_i^n \beta_i \mathbf{x}_i)^T \mathbf{x} - \mathbf{y}\right\|^2 + \alpha \left\|\beta_j(\sum_i^n \beta^T \mathbf{x}_i)\mathbf{x}\right\|^2 = \left\|(\mathbf{x}_i^T \mathbf{x}_j)_{ij}(\beta)_i - (\mathbf{y})_i\right\|^2 + \alpha(\beta)_i^T(\mathbf{x}_i^T \mathbf{x}_j)(\beta)_i$$

$$\|\mathbf{Kb} - \mathbf{y}\|^2 + \alpha \mathbf{b}^T \mathbf{Kb}$$

d) Es gilt:

$$J(\mathbf{b}) = \|\mathbf{Kb} - \mathbf{y}\|^2 + \alpha \mathbf{b}^T \mathbf{Kb}$$
$$= (\mathbf{Kb} - \mathbf{y})^T(\mathbf{Kb} - \mathbf{y}) + \alpha \mathbf{b}^T \mathbf{Kb}$$

Wir können $J(\mathbf{b})$ nach $\mathbf{b}$ ableiten (beachte $\mathbf{K} = \mathbf{K}^T$). Wenn die Ableitung gleich Null ist haben wir das Minimum:

$$\frac{\delta J(\mathbf{b})}{\delta \mathbf{b}} = 2\mathbf{K}(y - \mathbf{Kb}) + 2\alpha \mathbf{Kb} = 0$$

$$\Rightarrow \mathbf{b} = (\mathbf{K} + \alpha \mathbf{I})^{-1} \mathbf{y}$$

e) In der Formulierung in Aufgabe a) muss die Anzahl der Trainingsbeispiele größer gleich der Anzahl der Merkmale in einem Trainingsbeispiel sein, um $\mathbf{R}_{xx}$ berechnen zu können. Bei der Formulierung im Teil c) und d) wird eine Kernfunktion verwendet, die flexibel gewählt werden kann.

f)

1. Zunächst werden die Daten geladen. Dann wird die Matrix **K** und der Vektor **b** berechnet.

```
load('your_path\fish_fts.dat')

size_mond = size(Mond,1);
size_fts = size(Mond,1);
data = [Mond;Spitz];
size_data = size(data,1);

y = ones(size_data,1);
y(size_mond+1:size_data)=-1;

K = zeros(size_data,size_data);

for i = 1:size_data
  for j = 1:size_data
  K(i,j) = data(i,:) * data(j,:)';
  end
end

alpha = 0.1;
b = inv(K + alpha*eye(size_data,size_data))*y;
```

2. Es werden 8 Objekte falsch klassifiziert.

```
y_strich = sign(K*b);

fehler=0;

for i = 1:size_data
  if(y(i)~=y_strich(i))
  fehler = fehler + 1;
  end
end
```

3. Nach der Redefinition von **K** gibt es keine Fehlklassifikationen mehr.

```
K_neu = zeros(size_data,size_data);

d = 3;

for i = 1:size_data
  for j = 1:size_data
  K_neu(i,j) = (data(i,:) * data(j,:)' +  1)^d;
  end
end
```

```
alpha_neu = 0;
b_neu = inv(K_neu + alpha_neu*eye(size_data,size_data))*y;

y_strich_neu = sign(K_neu*b_neu);

fehler_neu=0;

for i = 1:size_data
  if(y(i)~=y_strich_neu(i))
  fehler_neu = fehler_neu + 1;
  end
end
```

Durch die die Redefinition von $\mathbf{K}$ werden nicht nur die paarweise Multiplikation zweier Vektoren sondern auch noch zusätzliche Monome betrachtet. Dadurch erreicht man eine viel größere Genauigkeit als bei der Klassifikation mit der Kernfunktion in 1. Zudem kann bei der neuen Kernfunktion $\alpha = 0$ gesetzt werden