言語処理のための機械学習入門

4.4. カーネル法
4.5. 対数線形モデル
4.6. 素性選択

永田 純平

カーネル法

$$\max. -\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y^{(i)}y^{(j)}x^{(i)}\cdot x^{(j)} + \sum_i \alpha_i$$

$$s.t. \sum_i \alpha_i y^{(i)} = 0$$

$$\alpha_i \geq 0; \forall_i$$

$$f(x) = \sum_i \alpha_i y^{(i)} x^{(i)} \cdot x - b$$

訓練データ$D = \{d^{(1)}, d^{(2)}, \cdots, d^{(|D|)}\}$は常に内積の形である

カーネル関数

事例の対を引数として高次元の内積を与える関数

引数である事例間の類似度をあらわす

$$\max. -\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y^{(i)}y^{(j)}K(d^{(i)},d^{(j)})+\sum_i \alpha_i$$

$$s.t. \sum_i \alpha_i y^{(i)} = 0$$

$$\alpha_i \geq 0; \forall_i$$

$$f(d)=\sum_i \alpha_i y^{(i)}K(d^{(i)},d)-b$$

- 木構造カーネル
- 文字列カーネル

- 多項式カーネル

$$K_{poly}\left(x^{(i)}, x^{(j)}\right) = (x^{(i)}x^{(j)} + r)^d$$

- RBFカーネル

$$K_{RBF}\left(x^{(i)}, x^{(j)}\right) = \exp(-s\left|x^{(i)} - x^{(j)}\right|^2)$$

対数線形モデル

- 確率的な分類器で、事例があるラベルをもつ確率を得る。

$$P(y|d) = \frac{1}{Z_{d,w}} \exp(w \cdot \phi(d,y))$$

※ $Z_{d,w} = \sum_y \exp(w \cdot \phi(d,y))$

$$y^* = \arg\max \frac{1}{Z_{d,w}} \exp(w \cdot \phi(d,y))$$

$$y^* = \arg\max w \cdot \phi(d,y)$$

※対数線形モデルでは

$P(d), P(d|y), P(d,y)$は計算できない。

対数線形モデルの学習

$$L(w) = \sum_{(d^{(i)}, y^{(i)}) \in D} \log P(y^{(i)} | d^{(i)}) - \frac{C}{2} |w|^2$$

上記の目的関数を最大化するが、この問題は解析的には求めることができない。

$$w^{new} = w^{old} + \epsilon \nabla_w L\left(w^{old}\right)$$

$$\nabla_w L(w_t) = \sum_{(d^{(i)}, y^{(i)}) \in D} \left(\phi\left(d^{(i)}, y^{(i)}\right) - \sum_y P\left(y \middle| d^{(i)}\right) \phi(d^{(i)}, y)\right) - Cw$$

素性選択

- ラベル付き訓練データと素性集合が与えられたときに自動的に有効な素性を選択する。

- 素性選択することで、分類の精度の向上やよりコンパクトなモデルにすることができる

自己相互情報量

$$PMI(w, c) = \log \frac{P(X_w = 1, C = c)}{P(X_w = 1) \times P(C = c)}$$

素性選択するためにクラスに依存しない指標が必要である

$$I_{average}(w) = \sum_{c} P(c) PMI(w, c)$$

$$I_{max}(w) = \max P(c) PMI(w, c)$$

情報利得

- 「ある素性が出現したか否か」という情報がエントロピーを減少させるかを素性の有効性とする

$$H(C) = -\sum_{c} P(c) \log P(c)$$

$$IG(w)$$
$$= H(C) - (P(X_w = 1)H(C|X_w = 1)$$
$$+ (P(X_w = 0)H(C|X_w = 0))$$