４．４　カーネル法
４．５　対数線形モデル
４．６　素性選択

08T4067L　横田　翔

４．４　カーネル法

カーネル法 ― 内積だけを用いて、学習や分類を行うこと

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。事例間の類似度を与える
(一般にKで表わされる)

カーネル関数の具体的な使用例

・木構造カーネル

－　言語処理において重要。木構造カーネルの値を計算するには、それぞれの木を、その木の中でのあらゆる部分木の頻度を素性としてベクトル化し、内積を取る。
　ただし、実際にベクトル化することは、計算量の点から現実的ではない。そこで、動的計画法を用い実際にベクトル化することなくカーネルの値を計算する。

・文字列カーネル

－　文字列をその部分文字列の頻度を素性としてベクトル化し、その内積を取る。木構造カーネル同様、動的計画法により、計算の高速化が可能。

・多項式カーネル(polynomial kernel)

－ すでにベクトル化されたものを入力として、それらに演算を施す。

例：K_{poly}で、二つのベクトル$x^{(i)}$と$x^{(j)}$に対して、次のような値を与える。

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

※rが学習結果に大きな影響を与えることは少なく
大抵の場合、r=1などと固定される。

d=2の場合は2次の多項式カーネル、d=3で3次の多項式カーネルなどと呼ぶ。

多項式カーネルのようなカーネルは、となどのベクトルを別の（高次元の）ベクトルに変換しているのと等価である。

- 動径素底関数カーネル
  - すでにベクトル化されたものを入力として、それらのベクトルに演算を施すような種類のカーネル関数。RBFカーネルとも呼ばれ次のように定義されている。

【定義】

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

※sは正の定数

４．５　対数線形モデル

対数線形モデル

- 確率的な分類機で、事例があるラベルを持つ確率を教えてくれる。最大エントロピーモデルとも呼ばれる。

４．５．１　素性表現の拡張と対数線形モデルの導入

素性に対する考え方の変更

今まで 事例が一つあればそれに対して一つの素性ベクトルが存在

ここでは 事例とラベルの組に対して素性が定義される

素性関数ϕk_kを定義し、これは事例dとラベルyを引数にとって、素性値$\phi_k(d,y)$を返すものとする。

$\phi_k(d,y)$を並べた$\phi(d,y)$が素性ベクトルとなる。

４．５．２　対数線形モデルの学習

以下のようなときの条件付き確率P(y|d)の対数尤度を考える

・訓練データDが与えられる

・条件付き確率P(y|d)以下のように表わされる

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

・$Z_{d,w}$は $\sum_y P(y|d)=1$ を保証する係数で

$$Z_{d,w}=\sum_y \exp(w\cdot\phi(d,y))$$と定義される

・wはパラメータである

$$\log P_{cond}(D) = \sum_{(d^{(i)}, y^{(i)}) \in D} \log P(y^{(i)} \mid d^{(i)})$$

$$= \sum_{d^{(i)}, y^{(i)} \in D} (\log \exp(w \cdot \phi(d^{(i)}, y^{(i)}) - \log Z_{d(i),w})$$

$$= \sum_{d^{(i)}, y^{(i)} \in D} (w \cdot \phi(d^{(i)}, y^{(i)} - \log Z_{d^{(i)},w}).$$

これで最大化しようとすると、あるクラスに出現した素性に対する重みだけ無限大になってしまってうまくいかない。

そこで目的関数に－(C/2)|w|^{2}を加えることで|w|が大きくなり過ぎないようにする。（Cは正の定数）

これを【正則化】という。

４．６　素性選択

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時に、自動的にそこから有効な素性を選択することを考えること。

４．６．１　自己相互情報量

自己相互情報量は以下のように表わされる。

$$PMI(w,c) = \log \frac{(w\text{と}c\text{の同時確率})}{(w\text{の確率}) \times (\text{ c の確率})}$$

これを詳しく表わすと

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

PMI(w,c)は単語wとクラスcがどのくらい共起しやすいかを示す。素性選択をするためにはクラスに依存しない指標が必要なので次の式を用いる。

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

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

この値が大きい素性が有効であるとされる。

定義の方法はこの他にも存在する。
いずれかの素性を値とする確率変数Wを考え

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

とすることもできる。

４．６．２　情報利得

情報利得
- 「ある素性が出現したか否か」という情報が、クラスに関する曖昧さ（エントロピー）をどれくらい減少させるかを表わしたもの

　クラスを表わす確率変数をCとすると、エントロピーH(C)は以下のように表わされる。

$$H(C) = -\sum P(c) \log P(c)$$

「ある素性が出現した」ことが分かっている場合の、条件付きエントロピーは以下のように表わされる。

$$H(C \mid X_w = 1) = -\sum_c P(c \mid X_w = 1) \log P(c \mid X_w = 1)$$

「ある素性が出現しなかった」ことが分かっている場合の、条件付きエントロピーは以下のように表わされる。

$$H(C \mid X_w = 0) = -\sum_c P(c \mid X_w = 0) \log P(c \mid X_w = 0)$$

平均的には

$$P(X_w = 1)H(C \mid X_w = 1) + P(X_w = 0)H(C \mid X_w = 0)$$

となる。

　素性wの情報利得IG(w)は、「エントロピーの平均的な減少量」として、次のように定義される。

$$IG(w) = H(C) - (P(Xw = 1)H(C \mid Xw = 0) + P(Xw = 0)H(C \mid Xw = 0)).$$