パターン認識と機械学習
第4章 線形識別モデル

橋本龍二

January 23, 2021

分類問題に関する3つのアプローチ：

1. 識別関数
   入力ベクトル $\boldsymbol{x}$ から直接クラスを推定する関する $f(\boldsymbol{x})$ を見つける。2クラスなら $f(\cdot)$ は2値を出力する。

2. 推論段階で条件付き確率 $p(\mathcal{C}_k|\boldsymbol{x})$ をモデル化したのち、$p(\mathcal{C}_k|\boldsymbol{x})$ を利用して決定理論を適用する。決定理論は、例えば

$$\min_j \sum_k L_{kj} p(\mathcal{C}_k|\boldsymbol{x})$$

   となるように j を決める期待損失最小化や、

$$\begin{aligned} \max \sum_k p(\boldsymbol{x} \in \mathcal{R}_k, C_k) &= \max \sum_k \int_{\mathcal{R}_k} p(\boldsymbol{x}, \mathcal{C}_k) d\boldsymbol{x} \\ &= \max \sum_k \int_{\mathcal{R}_k} p(\boldsymbol{x}) p(\mathcal{C}_k|\boldsymbol{x}) d\boldsymbol{x} \end{aligned}$$

   となるように決定領域 $\mathcal{R}_k$ を決める誤識別率最小化がある。
   推論段階において、$p(\mathcal{C}_k|\boldsymbol{x})$ を決める方法には次の二通りがある。

   (a) パラメトリックに分布を表現する。(確率的識別モデル)
   例えば、パラメータ $\boldsymbol{w}$ に分布を導入し、

$$y(\boldsymbol{x}) = f({}^T \boldsymbol{w}\boldsymbol{x} + w_0)$$

   から各クラスの事後確率を並べたベクトルを求める。

(b) クラスの事前確率 $p(\mathcal{C}_k)$ とクラスで条件づけられた確率密度 $p(\boldsymbol{x}|\mathcal{C}_k)$ から各クラスの事後確率：

$$p(\mathcal{C}_k|\boldsymbol{x}) = \frac{p(\mathcal{C}_k)p(\boldsymbol{x}|\mathcal{C}_k)}{p(\boldsymbol{x})}$$

を求める。(確率的生成モデル)

以上の 1、2(a)、2(b) の方法について詳しく見ていく。

1 識別関数

1.1 2 クラス

D 次元の入力ベクトル $\boldsymbol{x}$ が与えられた時、線形識別関数

$$y(\boldsymbol{x}) = {}^T\boldsymbol{w}\boldsymbol{x} + w_0$$

が 0 以上であればクラス $\mathcal{C}_1$ に割り当てることにする (決定境界：$y(\boldsymbol{x}) = 0$ を満たす点 $\boldsymbol{x}$ の集合)。

$${}^T\boldsymbol{w}\boldsymbol{x} + w_0 = 0$$

を満たす決定境界は $D-1$ 次元の超平面である ($D-1$ 個の値が決まれば残りは方程式より決定する)。
決定境界上の任意の 2 点 $\boldsymbol{x}_A, \boldsymbol{x}_B$ を考える。

$${}^T\boldsymbol{w}(\boldsymbol{x}_A - \boldsymbol{x}_B) = 0 \tag{1}$$

より、ベクトル $\boldsymbol{w}$ は決定面上の任意のベクトルと直交する。(w は平面の法線ベクトル、みたいなこと)
今、決定面上の任意の点 $\boldsymbol{x}$ をとり、原点から決定面への垂線をひく。この垂線は $\boldsymbol{w}$ と平行なので、定数 t を用いて $t\boldsymbol{w}$ と表すことにする。$t\boldsymbol{w}$ の長さが原点から決定面への距離である。(1) より

$$\begin{aligned}(\boldsymbol{x} - t\boldsymbol{w}, \boldsymbol{w}) &= 0 \\ (\boldsymbol{x}, \boldsymbol{w}) - t\|\boldsymbol{w}\|^2 &= 0 \\ \therefore t &= \frac{(\boldsymbol{x}, \boldsymbol{w})}{\|\boldsymbol{w}\|^2}\end{aligned}$$

したがって、原点から決定面の距離は、

$$\begin{aligned}\|t\boldsymbol{w}\| = \|\frac{(\boldsymbol{x}, \boldsymbol{w})}{\|\boldsymbol{w}\|^2}\boldsymbol{w}\| &= \frac{(\boldsymbol{x}, \boldsymbol{w})}{\|\boldsymbol{w}\|} \\ &= -\frac{w_0}{\|\boldsymbol{w}\|}\end{aligned}$$

(ただし、$\boldsymbol{w}$ がどちら向きでも r は変わらない訳だから厳密に距離を求めたくなったら絶対値を取る必要がある。)
次に、任意の点 $\boldsymbol{x}$ と決定面との距離を求める。$\boldsymbol{x}$ から決定面へ下ろした垂線と決定面との交点を $\boldsymbol{x}_\perp$($\boldsymbol{x}$ の決定面への直交射影) と置くと、垂線は $\boldsymbol{w}$ と平行なので、定数 s を用いて $s\boldsymbol{w}$ と表すことにする。すると $\boldsymbol{x}$ は、

$$\begin{aligned}
\boldsymbol{x} &= \boldsymbol{x}_\perp + s\boldsymbol{w} \\
&= \boldsymbol{x}_\perp + \frac{r}{\|\boldsymbol{w}\|}\boldsymbol{w} \\
&\quad (r = \|\boldsymbol{w}\|s)
\end{aligned} \tag{2}$$

と書ける。r が $\boldsymbol{x}$ と決定面の距離である。今、(2) の両辺に ${}^T\boldsymbol{w}$ を掛け w_0 を加えることで、

$$\begin{aligned}
{}^T\boldsymbol{w}\boldsymbol{x} + w_0 &= {}^T\boldsymbol{w}(\boldsymbol{x}_\perp + r\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}) + w_0 \\
&= {}^T\boldsymbol{w}\boldsymbol{x}_\perp + w_0 + r\|\boldsymbol{w}\| \\
\therefore y(\boldsymbol{x}) &= r\|\boldsymbol{w}\| \\
\therefore r &= \frac{y(\boldsymbol{x})}{\|\boldsymbol{w}\|}
\end{aligned}$$

まとめ

$$y(\boldsymbol{x}) = {}^T\boldsymbol{w}\boldsymbol{x} + w_0$$

なる線形識別関数において、

- $\boldsymbol{w}$ は決定面と直交する。すなわち $\boldsymbol{w}$ は決定面の方向を決定する。
- 原点から決定面までの距離は $-\frac{w_0}{\|\boldsymbol{w}\|}$ である。すなわち w_0 は決定面の位置を決定する。
- 任意の点 $\boldsymbol{x}$ と決定面との距離は $\frac{y(\boldsymbol{x})}{\|\boldsymbol{w}\|}$ である。すなわち $y(\boldsymbol{x})$ は決定面からの点 $\boldsymbol{x}$ の距離を与える。

1.2 多クラス

D 次元入力ベクトル $\boldsymbol{x}$ に対し、

$$y_k(\boldsymbol{x}) = {}^T\boldsymbol{w}_k\boldsymbol{x} + w_{k0}$$

を定義し、全ての $j \neq k$ に対し $y_k(\boldsymbol{x}) > y_j(\boldsymbol{x})$ ならば $\boldsymbol{x}$ をクラス $\mathcal{C}_k$ に分類することを考える。この時、クラス $\mathcal{C}_k$ とクラス $\mathcal{C}_j$ の決定境界は $y_k(\boldsymbol{x}) = y_j(\boldsymbol{x})$ である。すなわち、

$$^T(\boldsymbol{w}_k - \boldsymbol{w}_j)\boldsymbol{x} = -(w_{k0} - w_{j0})$$

なる $D-1$ 次元超平面。
識別器 $y_k(\boldsymbol{x})$ の決定領域は必ず凸領域である。つまり、任意の $\boldsymbol{x}_A, \boldsymbol{x}_B \in \mathcal{R}_k$ を取れば、線分 $\boldsymbol{x}_A\boldsymbol{x}_B$ 上の任意の点 $\hat{\boldsymbol{x}}$ は必ず $\mathcal{R}_k$ に属する。なぜなら、

$$\hat{\boldsymbol{x}} = \lambda\boldsymbol{x}_A + (1-\lambda)\boldsymbol{x}_B, \quad 0 \leq \lambda \leq 1$$

とすれば、

$$y_k(\hat{\boldsymbol{x}}) = \lambda y_k(\boldsymbol{x}_A) + (1-\lambda)y_k(\boldsymbol{x}_B)$$

今、任意の $j \neq k$ について $y_k(\boldsymbol{x}_A) > y_j(\boldsymbol{x}_A),\ y_k(\boldsymbol{x}_B) > y_j(\boldsymbol{x}_B)$ が成り立つのだから、$\hat{\boldsymbol{x}} \in \mathcal{R}_k$

2　確率的生成モデル

3　確率的識別モデル