

# PRML 輪読会

## 決定理論 (1.5.1 節 – 1.5.3 節)

---

鈴木拓己

November 20, 2024

決定理論の枠組み

誤識別率の最小化

期待損失の最小化

棄却オプション

演習問題 (Appendix)

決定理論の枠組み

誤識別率の最小化

期待損失の最小化

棄却オプション

演習問題 (Appendix)

決定理論は、確率論と組み合わせることによって、パターン認識で遭遇する、不確かさを含む状況における最適な意思決定を行うことを可能とする。

### 問題設定

- $x$ : 入力ベクトル,  $t$ : 目標変数
  - 回帰問題:  $t$  は連続変数
  - 分類問題:  $t$  はクラスラベル
- 確率分布  $p(x, t)$ : 入力ベクトルと目標変数に関する不確実性を完全に記述するものであり、訓練データ集合から推論する。

実際の応用では  $t$  の特定の値を予測したり、あるいはより一般に  $t$  が取り得る値に応じて特定の行動を取ることが多いが、決定理論ではこうした側面を扱う。

### 例: 医療診断問題

問題: 患者の X 線画像を取り, その患者が癌であるかどうかを判定したい.

- 入力ベクトル  $x$ : 画像のピクセル強度の集合
- 出力変数:  $t = 0$  (癌であるクラス  $C_1$ ) or  $t = 1$  (癌でないクラス  $C_2$ )

目標: 新たな患者の X 線画像  $x$  が与えられたとき, 画像に 2 つのクラスのどちらかを割り当てること.

このとき, 同時分布  $p(x, t)$  を推論することにより, 状況の完全な確率的記述ができるが, 最終的に患者に治療を施すかどうかを決めるために, ある適当な規準の上で最適な決定をする必要がある.

与えられた画像に対し、2つのクラスの確率をそれぞれ  $p(C_k | \mathbf{x})$  とすると、これらの確率はベイズの定理より、

$$p(C_k | \mathbf{x}) = \frac{p(\mathbf{x} | C_k)p(C_k)}{p(\mathbf{x})}.$$

ここに現れる量はすべて同時分布  $p(C_k, \mathbf{x})$  から周辺化あるいは適当な変数に関する条件付きの操作によって得られる。

$p(C_k)$  をクラス  $C_k$  の事前確率、 $p(C_k | \mathbf{x})$  を対応する事後確率とする。

- すなわち、 $p(C_1)$  は X 線計測をする前に人間が癌である確率であり、 $p(C_1 | \mathbf{x})$  は X 線画像に含まれる情報を得たことによって、ベイズの定理を用いて修正された確率。

誤ったクラスに分類する可能性を最小にしたければ、直観的には、高い事後確率を持つクラスを選べばよい。

決定理論の枠組み

誤識別率の最小化

期待損失の最小化

棄却オプション

演習問題 (Appendix)

まず、できるだけ誤識別を少なくしたいということだけを目標とすることを考える。

決定のためには、 $x$  の各値に対してクラスを 1 つ割り当てるための規則が必要であり、これは入力空間を決定領域によって分割すること。

- 決定領域  $\mathcal{R}_k$ : クラス  $\mathcal{C}_k$  に対応する入力空間の部分集合
- 決定境界: 決定領域の間の境界

各決定領域は連続とは限らず、いくつかの領域に分かれることもあり得る。

入力空間の各点  $x$  にどのクラスを割り振るかの決定規則はどのように選んでもよい。



医療診断問題の例における 2 クラス分類問題の場合、誤識別が起きる確率は、

$$\begin{aligned} p(\text{誤り}) &= p(\boldsymbol{x} \in \mathcal{R}_1, \mathcal{C}_2) + p(\boldsymbol{x} \in \mathcal{R}_2, \mathcal{C}_1) \\ &= \int_{\mathcal{R}_1} p(\boldsymbol{x}, \mathcal{C}_2) d\boldsymbol{x} + \int_{\mathcal{R}_2} p(\boldsymbol{x}, \mathcal{C}_1) d\boldsymbol{x} \\ &= \int_{\mathcal{R}_1} p(\mathcal{C}_2 \mid \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} + \int_{\mathcal{R}_2} p(\mathcal{C}_1 \mid \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}. \end{aligned} \tag{2.1}$$

- $p(\text{誤り})$  の最小化  $\iff$  (2.1) の積分値が小さくなるようなクラスの割り振り
- つまり,  $p(\boldsymbol{x}, \mathcal{C}_1) > p(\boldsymbol{x}, \mathcal{C}_2)$  なら,  $\boldsymbol{x}$  にはクラス  $\mathcal{C}_1$  を割り当てるべき
- $p(\boldsymbol{x})$  はクラスに依らず共通なので,  $p(\text{誤り})$  の最小化は, 事後確率  $p(\mathcal{C}_k \mid \boldsymbol{x})$  が最大のクラスへの割り当てと等価

より一般の  $K$  クラスの場合は、以下で与えられる正解の確率を最大化する方が易しい。

$$\begin{aligned} p(\text{正解}) &= \sum_{k=1}^K p(\mathbf{x} \in \mathcal{R}_k, \mathcal{C}_k) = \sum_{k=1}^K \int_{\mathcal{R}_k} p(\mathbf{x}, \mathcal{C}_k) d\mathbf{x} \\ &= \sum_{k=1}^K \int_{\mathcal{R}_k} p(\mathcal{C}_k \mid \mathbf{x}) p(\mathbf{x}) d\mathbf{x}. \end{aligned} \tag{2.2}$$

- $p(\text{正解})$  の最大化  $\iff$  (2.2) の積分値が大きくなるようなクラスの割り振り
- これは各  $\mathbf{x}$  を  $p(\mathbf{x}, \mathcal{C}_k)$  が最大となるクラスに割り当てるように決定領域  $\mathcal{R}_k$  を選んだとき最大となる
- $p(\mathbf{x})$  はクラスに依らず共通なので、 $p(\text{正解})$  の最大化は、事後確率  $p(\mathcal{C}_k \mid \mathbf{x})$  が最大のクラスへの割り当てと等価

## 誤識別率の最小化: 1次元の入力からなる2クラス分類問題の場合

右図のように決定規則を定めると,

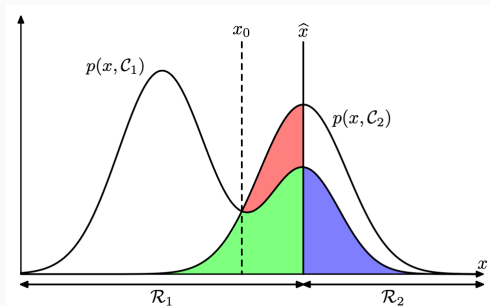
- $x \geq \hat{x} \implies x$  はクラス  $C_2$  に分類
- $x < \hat{x} \implies x$  はクラス  $C_1$  に分類

となる. このとき,

- $x \geq \hat{x}$  における誤差  $\implies$  青の領域
- $x < \hat{x}$  における誤差  $\implies$  (赤 + 緑) の領域

であり, 決定境界  $\hat{x}$  の位置を動かしても (青 + 緑) の領域は変わらず, 赤の領域のみ変わるので, 最適な決定境界は  $\hat{x} = x_0$  (両曲線の交点) とわかる.

これは, 誤識別率最小の決定規則と等価 (入力  $x$  の各値を事後確率  $p(C_k | x)$  の大きい方に分類).



**Figure 1:**  $x$  に対する, 2つのクラスの同時分布  $p(x, C_k)$  と決定境界  $x = \hat{x}$  の模式図

決定理論の枠組み

誤識別率の最小化

期待損失の最小化

棄却オプション

演習問題 (Appendix)

誤識別率の最小化は、分類問題の決定を完全に解決するわけではない。

例: 医療診断問題 (再訪)

- 癌でない患者を誤って癌と診断した場合、さらに検査が必要になる程度で済むが、癌の患者を健康であると診断した場合、治療が足りず、患者が死に至る可能性がある。
- 明らかに、前者の誤りを増やすことになったとしても、後者の誤りを減らすべき。

このような問題は、損失関数の最小化 (  $\iff$  効用関数の最大化 ) によって定式化可能。

入力  $x$  に対する真のクラスを  $C_k$  とし,  $x$  を  $C_j$  に割り当てたとしたとき, こうむる損失の値を  $L_{kj}$  で表すと, それを  $(k, j)$ -成分とする損失行列  $L$  を考えることができる. 例えば, 癌の例で言えば, 損失行列は以下のような形になる:

$$L = \begin{pmatrix} 0 & 1000 \\ 1 & 0 \end{pmatrix}. \quad \left( k = \begin{cases} 1, & \text{(癌)} \\ 2 & \text{(正常)} \end{cases}, j = \begin{cases} 1, & \text{(癌)} \\ 2 & \text{(正常)} \end{cases} \right)$$

上記の例では,

- 正しい決定 ( $k = j = 1, 2$ ) をすれば損失は無く,
- 健康な患者を癌と診断したとき ( $k = 2, j = 1$ ) の損失を 1 とし,
- 癌の患者を健康と診断したとき ( $k = 1, j = 2$ ) の損失を 1000

としている.

## 期待損失の最小化

損失関数を最小にするのが最適解だが、損失関数は未知である真のクラスに依存。

したがって、与えられた入力ベクトル  $\boldsymbol{x}$  に対して、真のクラスの不確実性  $p(\boldsymbol{x}, \mathcal{C}_k)$  を用いて、以下で計算される損失の平均 (期待損失) を最小化。

$$\begin{aligned}\mathbb{E}[L] &= \sum_k \sum_j \int_{\mathcal{R}_j} L_{kj} p(\boldsymbol{x}, \mathcal{C}_k) d\boldsymbol{x} = \sum_j \int_{\mathcal{R}_j} \sum_k L_{kj} p(\boldsymbol{x}, \mathcal{C}_k) d\boldsymbol{x} \\ &= \sum_j \int_{\mathcal{R}_j} \sum_k L_{kj} p(\mathcal{C}_k \mid \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}.\end{aligned}$$

各  $\boldsymbol{x}$  は決定領域  $\mathcal{R}_j$  のどれかに独立に割り当てることができる。

したがって、事後クラス確率  $p(\mathcal{C}_k \mid \boldsymbol{x})$  がわかれば、期待損失を最小化する決定規則は、新たな  $\boldsymbol{x}$  を  $\sum_k L_{kj} p(\mathcal{C}_k \mid \boldsymbol{x})$  が最小になるようなクラス  $j$  に割り当てることである。

決定理論の枠組み

誤識別率の最小化

期待損失の最小化

棄却オプション

演習問題 (Appendix)



クラス分類の誤差が起きるのは、入力空間の事後確率  $p(\mathcal{C}_k | \boldsymbol{x})$  の最大値が 1 よりも大幅に小さい (同時確率  $p(\boldsymbol{x}, \mathcal{C}_k)$  の値が拮抗している) ときであり、こうした領域は入力  $\boldsymbol{x}$  がどのクラスに属するかが不確か。

応用によっては、このような場合には無理に決定をせず、クラス分類の決定を行ったものに対する誤差を小さくする方がよい (棄却オプション)。

### 例: 医療診断問題 (再訪)

- 正しいクラスに分類できるかどうかをはっきりしている X 線画像は、自動的に分類するシステムを用いることが適当。
- 正しいクラスに分類できるかどうかが曖昧な場合については、専門家の判断にゆだねる (決定の棄却)。

## 棄却オプション

棄却オプションは、閾値  $\theta$  を導入し、事後確率  $p(C_k | x)$  の最大値が  $\theta$  以下となるような入力  $x$  に対しては決定を棄却することによって実現可能.

- 右図では、2つの事後確率の大きい方が閾値  $\theta$  以下の場合に入力  $x$  を棄却
- $K$  クラス分類問題においては、閾値  $\theta$  は
- $\theta = 1 \implies$  すべての事例が棄却
  - $\theta < 1/K \implies$  どの事例も棄却されない
- となる ( $\because \sum_k p(C_k | x) = 1$ ).
- このように、 $\theta$  の値によって棄却される事例の割合が制御される.

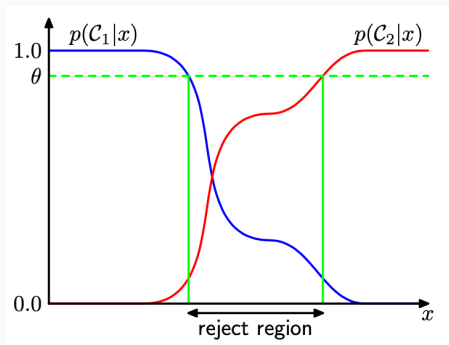


Figure 2: 棄却オプションの図

棄却決定がなされたときの損失を考慮した損失行列を与えれば、期待損失を最小にするように棄却の規準を容易に一般化することができる (演習 1.24, Appendix 参照).

### 演習 1.24 (標準) www

クラス分類問題を考え、クラス  $C_k$  からの入力ベクトルをクラス  $C_j$  と分類したときの損失行列を  $L_{kj}$  とし、棄却オプションを選んだときの損失を  $\lambda$  とする. このとき期待損失を最小とする決定規準を見つけよ. 損失行列が  $L_{kj} = 1 - I_{kj}$  のときは、決定規準は 1.5.3 節で議論した棄却規準に帰着されることを確かめよ. また、 $\lambda$  と棄却閾値  $\theta$  にはどんな関係があるか?

決定理論の枠組み

誤識別率の最小化

期待損失の最小化

棄却オプション

演習問題 (Appendix)

### 演習 1.21 (標準)

2 つの非負の数  $a$  と  $b$  があったとき,  $a \leq b$  なら  $a \leq (ab)^{1/2}$  であることを示せ. この結果を使って, 2 クラスの分類問題の決定領域を誤識別率が最小になるように選ぶと, この確率が

$$p(\text{誤り}) \leq \int \{p(\mathbf{x}, C_1)p(\mathbf{x}, C_2)\}^{1/2} d\mathbf{x}$$

を満たすことを示せ.

(解答)  $a, b \geq 0$  より,  $a \leq b \implies a^{1/2} \leq b^{1/2} \implies a \leq (ab)^{1/2}$  である. 誤識別率が最小となる決定規則は,  $p(\mathbf{x}, C_2) \leq p(\mathbf{x}, C_1)$  (resp.  $p(\mathbf{x}, C_1) \leq p(\mathbf{x}, C_2)$ ) なる  $\mathbf{x}$  の決定領域を  $\mathcal{R}_1$  (resp.  $\mathcal{R}_2$ ) と選べばよいので,

$$p(\mathbf{x}, C_2) \leq p(\mathbf{x}, C_1) \implies \int_{\mathcal{R}_1} p(\mathbf{x}, C_2) d\mathbf{x} \leq \int_{\mathcal{R}_1} \{p(\mathbf{x}, C_1)p(\mathbf{x}, C_2)\}^{1/2} d\mathbf{x},$$

$$p(\mathbf{x}, C_1) \leq p(\mathbf{x}, C_2) \implies \int_{\mathcal{R}_2} p(\mathbf{x}, C_1) d\mathbf{x} \leq \int_{\mathcal{R}_2} \{p(\mathbf{x}, C_1)p(\mathbf{x}, C_2)\}^{1/2} d\mathbf{x},$$

ゆえ,

$$p(\text{誤り}) = \int_{\mathcal{R}_1} p(\mathbf{x}, C_2) d\mathbf{x} + \int_{\mathcal{R}_2} p(\mathbf{x}, C_1) d\mathbf{x} \leq \int \{p(\mathbf{x}, C_1)p(\mathbf{x}, C_2)\}^{1/2} d\mathbf{x}. \quad \square$$

### 演習 1.22 (基本) www

$L_{kj}$  を要素とする損失行列が与えられたとき、期待リスクが最小になるのは、各  $\mathbf{x}$  に対し、 $\sum_k L_{kj} p(C_k | \mathbf{x})$  を最小にするクラスを選んだときである。損失行列が  $L_{kj} = 1 - I_{kj}$  で与えられたとき、これが最大事後確率のクラスを選ぶ規準に帰着されることを確かめよ。ただし、 $I_{kj}$  は単位行列の成分を表す。また、この損失行列はどのように解釈できるか？

(解答) 与えられた損失行列より、

$$\begin{aligned}\sum_k L_{kj} p(C_k | \mathbf{x}) &= \sum_k (1 - I_{kj}) p(C_k | \mathbf{x}) = \sum_k p(C_k | \mathbf{x}) - \sum_k I_{kj} p(C_k | \mathbf{x}) \\ &= 1 - \sum_k I_{kj} p(C_k | \mathbf{x}) = 1 - p(C_j | \mathbf{x})\end{aligned}$$

を得る。期待損失が最小になるのは、各  $\mathbf{x}$  に対し、 $\sum_k L_{kj} p(C_k | \mathbf{x})$  を最小にするクラスを選んだときであり、これは  $p(C_j | \mathbf{x})$  を最大にするクラスを選ぶことと同値であるので、最大事後確率のクラスを選ぶ規準に帰着される。また、損失行列  $L_{kj} = 1 - I_{kj}$  は対角成分が 0 で、対角成分以外の成分が 1 であるような行列なので、期待損失の最小化は、誤識別率の最小化と等価である。

### 演習 1.23 (基本)

一般の場合に、損失行列とクラスに対する事前確率が与えられたときに、期待損失を最小にする規準を導け。

(解答) 期待損失を最小にする規準は、損失行列  $L$  と事後確率  $p(C_k | \mathbf{x})$  を用いて、 $\sum_k L_{kj} p(C_k | \mathbf{x})$  を最小にする規準と等価であった。ここで、事後確率  $p(C_k | \mathbf{x})$  は、ベイズの定理を用いると、クラスに対する事前確率  $p(C_k)$  を用いて、

$$p(C_k | \mathbf{x}) = \frac{p(\mathbf{x} | C_k)p(C_k)}{p(\mathbf{x})}$$

と書けるので、

$$\sum_k L_{kj} p(C_k | \mathbf{x}) = \frac{1}{p(\mathbf{x})} \sum_k L_{kj} p(\mathbf{x} | C_k) p(C_k)$$

を得る。したがって、

$$\operatorname{argmin}_j \sum_k L_{kj} p(C_k | \mathbf{x}) = \operatorname{argmin}_j \sum_k L_{kj} p(\mathbf{x} | C_k) p(C_k)$$

となる。

## 演習問題 (Appendix)

### 演習 1.24 (標準) www

クラス分類問題を考え、クラス  $C_k$  からの入力ベクトルをクラス  $C_j$  と分類したときの損失行列を  $L_{kj}$  とし、棄却オプションを選んだときの損失を  $\lambda$  とする．このとき期待損失を最小とする決定規準を見つけよ．損失行列が  $L_{kj} = 1 - I_{kj}$  のときは、決定規準は 1.5.3 節で議論した棄却規準に帰着されることを確かめよ．また、 $\lambda$  と棄却閾値  $\theta$  にはどんな関係があるか？

(解答) 棄却を考えない場合は、期待損失の最小化は  $\sum_k L_{kj} p(C_k | \mathbf{x})$  の最小化と等価なので、棄却オプションを考慮したときの期待損失を最小とする決定規準は、

$$\min_j \sum_k L_{kj} p(C_k | \mathbf{x}) < \lambda \implies \text{クラス } C_j \text{ を選択,}$$

$$\min_j \sum_k L_{kj} p(C_k | \mathbf{x}) \geq \lambda \implies \text{棄却}$$

となる．損失行列が  $L_{kj} = 1 - I_{kj}$  のときは、演習 1.22 より、期待損失の最小化は  $1 - p(C_j | \mathbf{x})$  の最小化となるので、期待損失を最小とする決定規準は、

$$1 - p(C_j | \mathbf{x}) < \lambda \iff p(C_j | \mathbf{x}) > 1 - \lambda \implies \text{クラス } C_j \text{ を選択,}$$

$$1 - p(C_j | \mathbf{x}) \geq \lambda \iff p(C_j | \mathbf{x}) \leq 1 - \lambda \implies \text{棄却}$$

となる．したがって、棄却閾値を  $\theta = 1 - \lambda$  としたとき、決定規準は 1.5.3 節で議論した棄却規準に帰着される．