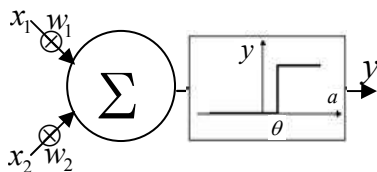


1. 下記の文章の空欄を埋めなさい。

- 古典的パターン認識理論では、入力 x に対してクラス ω_i への帰属度を表す度合いを各クラスについて計算し、その度合いが最も高いクラス ω_i に分類するという方法がよく用いられる。例えば、帰属度として事後確率 $P(\omega_i|x)$ を用いる場合はこの値の大小で識別を行うが、これは [] 基準のもとで [] を最小化する識別方法であることが知られている。
- この場合、事前生起確率 $P(\omega_i)$ と、クラス ω_i に属するパターン x が生起する確率 $P(x|\omega_i)$ から [] の定理を用いて事後確率 $P(\omega_i|x)$ を計算することになる。しかし、 $P(x|\omega_i)$ を確率の形式で求めることは困難であるため、[] で代替することが一般的に行われている。クラス数を N とした場合の事後確率は次式のように計算される。
- 一方、単純類似度法では、入力 x と各クラスを代表するパターンとの [] を計算し、その値が最も大きいクラスに分類するという計算が行われる。これを拡張した [] では、[] 行列の固有ベクトルから成る複数の代表パターンとの内積計算を行い、それを対応する固有値によって重み付けした値を類似度として用いるということが行われる。この計算は [] で行われている射影成分の大小で分類を行う方法と本質的に等価である。なぜなら、両者ともパターン間の [] に基づいて識別を行っているからである。
- これ以外に、[] で行われているように、入力に最も近い既知のパターン（プロトタイプ）が帰属するクラスに分類する方法もある。この場合には、各クラスへの帰属度は計算されない。この方法において、[] という手法を用いると、不要なプロトタイプを削除することができる。この際に残されるプロトタイプは、識別境界付近の誤識別を起こしやすい特殊例となる。
- このように、特殊例だけを記憶しそれによって識別を行う手法として [] がある。この方法ではマージン最大化基準に基づいて線形識別面を求めるために、[] を用いた最適化計算を行っているが、この計算によって未定係数 α が [] となるパターンは特殊例すなわち [] となる。
- この例以外にも [] では、誤識別をおこしたトレーニングデータの重みを増して識別器のトレーニングを行いながら、識別器の系列を発生させ、これらの線形結合によって識別器の出力を得ている。このことは、誤識別を起こしやすい部分の重みを増すという意味で、特殊例を重視した識別法であると言える。
- 以上のように、古典から現代的なパターン認識理論に至る世界観は「典型」から「特殊」へと変遷してきたと言える。しかし、特殊事例を用いた最初の識別法は、[] という最も古典的な識別法であったことは特筆に価する。

2. 次に示す閾値論理素子を用いて NAND 素子の学習を行う問題を考える。 $w = (0.1, 0)$ $\theta = -1.2$, $\alpha = 0.25$ として、下記の表を完成させなさい



	w1	w2	θ	x1	x2	a	y	t	$\alpha(t-y)$	$\delta w1$	$\delta w2$	$\delta \theta$
	0.1	0	-1.2	0	0	0	1	1	0	0	0	0
				0	1							
				1	0							
				1	1							
				0	0							
				0	1							
				1	0							
				1	1							
				0	0							
				0	1							
				1	0							
				1	1							

3. d 次元空間中に d 個の独立なベクトルが存在する. これらのベクトルを用いて共分散行列を求めたとき, そのランクはどれだけになるか?
4. 共分散行列 Σ の固有値と固有ベクトルがそれぞれ λ_i と $\boldsymbol{\varphi}_i$ ($i=1\cdots n$) であるとき, 共分散行列 Σ を λ_i と $\boldsymbol{\varphi}_i$ を用いて表しなさい.
5. d 次元ベクトル \boldsymbol{f}_i ($i=1,\dots,M$) が与えられている. その固有ベクトルを求める方法について考える. $F=[\boldsymbol{f}_1 \ \cdots \ \boldsymbol{f}_M]$ とすると, 共分散行列を M 倍した行列は $\sum_{i=1}^M \boldsymbol{f}_i \boldsymbol{f}_i^T = FF^T$ と表せる. この行列の固有ベクトルを FF^T ではなく $F^T F$ の固有ベクトルから求める方法と, その利点について述べなさい.
6. $f(x)$ の Fourier 変換を $F(\omega)=\int_{-\infty}^{\infty} f(x)e^{-j\omega x} dx$ とする. このとき, $f(x)$ が平行移動したパターン $g(x)=f(x-\mu)$ の Fourier 変換を $F(\omega)$ を用いた式で表しなさい. また, μ に対して不変な特徴を抽出する方法の一つとして, 自己相関関数 $R(\tau)=\int_{-\infty}^{\infty} f(x-\mu+\tau)f(x-\mu)dx$ が考えられる. $R(\tau)$ の Fourier 変換が $\|F(\omega)\|^2$ となることを示しなさい. また, この不変特徴の問題点について説明しなさい.
7. パターン \mathbf{x} を正規直交基底 \mathbf{u}_i ($i=1,\dots,n$) が張る部分空間に正射影する行列 P を $U=(\mathbf{u}_1 \ \cdots \ \mathbf{u}_n)$ という行列を用いて表現しなさい.

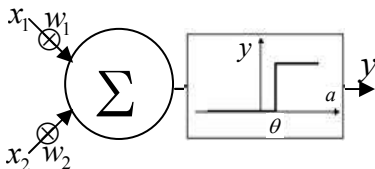
8. 7において、 C 個の部分空間が存在する場合、各部分空間への射影行列 P_i ($i=1, \dots, C$) を用いて、「 $\mathbf{x}^T P_i \mathbf{x}$ を最大化する i を求める問題」と、「 $\|\mathbf{x} - P_i \mathbf{x}\|$ を最小化する i を求める問題」は本質的に同じであることを示しなさい.
9. SVM の学習時に必要となる計算量・メモリ量がデータ数およびデータの次元数とどのような関係にあるかについて述べなさい.
10. 無限にトレーニングデータが与えられる場合の最近傍識別器の識別性能の限界について説明しなさい.
11. 最近傍探索における次元の呪縛とは何かについて述べなさい.
12. k-means アルゴリズムと EM アルゴリズムの類似性と相違性について述べよ.

13. Exclusive OR のパターンを学習することができる TLU ネットワークを図示し、その構造の必然性を説明しなさい。
-

1. 下記の文章の空欄を埋めなさい。(小さい四角は1問1点, 大きい四角は3点, 合計20点)
- 古典的パターン認識理論では, 入力 x に対してクラス ω_i への帰属度を表す度合いを各クラスについて計算し, その度合いが最も高いクラス ω_i に分類するという方法がよく用いられる. 例えば, 帰属度として事後確率 $P(\omega_i|x)$ を用いる場合はこの値の大小で識別を行うが, これは **0-1 損失** 基準のもとで **損失** を最小化する識別方法であることが知られている.
 - この場合, 事前生起確率 $P(\omega_i)$ と, クラス ω_i に属するパターン x が生起する確率 $P(x|\omega_i)$ から **Bayes** の定理を用いて事後確率 $P(\omega_i|x)$ を計算することになる. しかし, $P(x|\omega_i)$ を確率の形式で求めることは困難であるため, **確率密度関数 $p(x|\omega_i)$** で代替することが一般的に行われている. クラス数を N とした場合の事後確率は次式のように計算される.

$$P(\omega_i|x) = \frac{p(x|\omega_i) P(\omega_i)}{\sum_{j=1}^N P(\omega_j) p(x|\omega_j)}$$

- 一方, 単純類似度法では, 入力 x と各クラスを代表するパターンとの **余弦** を計算し, その値が最も大きいクラスに分類するという計算が行われる. これを拡張した **複合類似度法** では, **自己相関行列** の固有ベクトルから成る複数の代表パターンとの内積計算を行い, それを対応する固有値によって重み付けした値を類似度として用いるということが行われる. この計算は **部分空間法** で行われている射影成分の大小で分類を行う方法と本質的に等価である. なぜなら, 両者ともパターン間の **角度** に基づいて識別を行っているからである.
 - これ以外に, **最近傍識別器** で行われているように, 入力に最も近い既知のパターン (プロトタイプ) が帰属するクラスに分類する方法もある. この場合には, 各クラスへの帰属度は計算されない. この方法において, **Voronoi Condensing** という手法を用いると, 不要なプロトタイプを削除することができる. この際に残されるプロトタイプは, 識別境界付近の誤識別を起ししやすい特殊例となる.
 - このように, 特殊例だけを記憶しそれによって識別を行う手法として **SVM** がある. この方法ではマージン最大化基準に基づいて線形識別面を求めるために, **ラグランジュの未定係数法** を用いた最適化計算を行っているが, この計算によって未定係数 α が **$\alpha \neq 0$** となるパターンは特殊例すなわち **サポートベクター** となる.
 - この例以外にも **ADA Boosting** では, 誤識別をおこしたトレーニングデータの重みを増して識別器のトレーニングを行いながら, 識別器の系列を発生させ, これらの線形結合によって識別器の出力を得ている. このことは, 誤識別を起ししやすい部分の重みを増すという意味で, 特殊例を重視した識別法であると言える.
 - 以上のように, 古典から現代的なパターン認識理論に至る世界観は「典型」から「特殊」へと変遷してきたと言える. しかし, 特殊事例を用いた最初の識別法は, **最近傍識別** という最も古典的な識別法であったことは特筆に価する.
2. 次に示す閾値論理素子を用いて NAND 素子の学習を行う問題を考える. $w = (0.1, 0)$ $\theta = -1.2$, $\alpha = 0.25$ として, 下記の表を完成させなさい



w1	w2	θ	x1	x2	a	y	t	$\alpha(t-y)$	$\delta w1$	$\delta w2$	$\delta \theta$
0.1	0	-1.2	0	0	0	1	1	0	0	0	0
0.1	0	-1.2	0	1	0	1	1	0	0	0	0
0.1	0	-1.2	1	0	0.1	1	1	0	0	0	0
0.1	0	-1.2	1	1	0.1	1	0	-0.25	-0.3	-0.3	0.25
-0.15	-0.25	-0.95	0	0	0	1	1	0	0	0	0
-0.15	-0.25	-0.95	0	1	-0	1	1	0	0	0	0
-0.15	-0.25	-0.95	1	0	-0	1	1	0	0	0	0
-0.15	-0.25	-0.95	1	1	-0	1	0	-0.25	-0.3	-0.3	0.25
-0.4	-0.5	-0.7	0	0	0	1	1	0	0	0	0
-0.4	-0.5	-0.7	0	1	-1	1	1	0	0	0	0
-0.4	-0.5	-0.7	1	0	-0	1	1	0	0	0	0
-0.4	-0.5	-0.7	1	1	-1	0	0	0	0	0	0

3. d 次元空間中に d 個の独立なベクトルが存在する. これらのベクトルを用いて共分散行列を求めたとき, そのランクはどれだけになるか?

$d-1$

4. 共分散行列 Σ の固有値と固有ベクトルがそれぞれ λ_i と $\boldsymbol{\varphi}_i$ ($i=1, \dots, n$) であるとき, 共分散行列 Σ を λ_i と $\boldsymbol{\varphi}_i$ を用いて表しなさい. (15 点)

$$\Sigma = V \Lambda V^T = \begin{bmatrix} \boldsymbol{\varphi}_1 & \cdots & \boldsymbol{\varphi}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\varphi}_1^T \\ \vdots \\ \boldsymbol{\varphi}_n^T \end{bmatrix} = \sum_{i=1}^n \lambda_i \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^T$$

5. d 次元ベクトル \boldsymbol{f}_i ($i=1, \dots, M$) が与えられている. その固有ベクトルを求める方法について考える. $F = [\boldsymbol{f}_1 \ \cdots \ \boldsymbol{f}_M]$ とすると, 共分散行列を M 倍した行列は $\sum_{i=1}^M \boldsymbol{f}_i \boldsymbol{f}_i^T = FF^T$ と表せる. この行列の固有ベクトルを FF^T ではなく $F^T F$ の固有ベクトルから求める方法と, その利点について述べなさい.

$F^T F$ の固有ベクトルを $\boldsymbol{\varphi}$, 固有値を λ とすると, $F^T F \boldsymbol{\varphi} = \lambda \boldsymbol{\varphi}$ が成立する. この式の左から F をかけると, $FF^T F \boldsymbol{\varphi} = \lambda F \boldsymbol{\varphi}$ が得られる. $FF^T (F \boldsymbol{\varphi}) = \lambda (F \boldsymbol{\varphi})$ と見なすと, FF^T の固有ベクトルは, $F \boldsymbol{\varphi}$ と表せる. この方法のメリットは, $d \ll M$ の場合, FF^T が $d \times d$ であるのに対し $F^T F$ は $M \times M$ と小さくなることである.

6. $f(x)$ の Fourier 変換を $F(\omega) = \int_{-\infty}^{\infty} f(x) e^{-j\omega x} dx$ とする. このとき, $f(x)$ が平行移動したパターン $g(x) = f(x - \mu)$ の Fourier 変換を $F(\omega)$ を用いた式で表しなさい. また, μ に対して不変な特徴を抽出する方法の一つとして, 自己相関関数 $R(\tau) = \int_{-\infty}^{\infty} f(x - \mu + \tau) f(x - \mu) dx$ が考えられる. $R(\tau)$ の Fourier 変換が $\|F(\omega)\|^2$ となることを示しなさい. また, この不変特徴の問題点について説明しなさい.

$$G(\omega) = \int_{-\infty}^{\infty} f(x - \mu) e^{-j\omega x} dx = \int_{-\infty}^{\infty} f(X) e^{-j\omega(X + \mu)} dX = e^{-j\omega\mu} \int_{-\infty}^{\infty} f(X) e^{-j\omega X} dX = e^{-j\omega\mu} F(\omega)$$

$$\int_{-\infty}^{\infty} R(\tau) e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x - \mu + \tau) f(x - \mu) dx e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(X + \tau) f(X) dX e^{-j\omega\tau} d\tau$$

変数変換を行い, 次の結果を得る.

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(T) f(X) dX e^{-j\omega(T-X)} dT = \int_{-\infty}^{\infty} f(T) e^{-j\omega T} dT \int_{-\infty}^{\infty} f(X) e^{j\omega X} dX = F(\omega) F^*(\omega) = \|F(\omega)\|^2$$

このようにして得られた不変特徴は, 各周波数成分間の位相情報まで消されているため, 識別に利用できる特徴が少なくなってしまうという問題点がある.

7. パターン \boldsymbol{x} を正規直交基底 \boldsymbol{u}_i ($i=1, \dots, n$) が張る部分空間に正射影する行列 P を $U = (\boldsymbol{u}_1 \ \cdots \ \boldsymbol{u}_n)$ という行列を用いて表現しなさい.

$$P = UU^T$$

8. 7において, C 個の部分空間が存在する場合, 各部分空間への射影行列 P_i ($i=1, \dots, C$) を用いて, 「 $\boldsymbol{x}^T P_i \boldsymbol{x}$ を最大化する i を求める問題」と, 「 $\|\boldsymbol{x} - P_i \boldsymbol{x}\|$ を最小化する i を求める問題」は本質的に同じであることを示しなさい.

$\|\boldsymbol{x} - P_i \boldsymbol{x}\|$ を最小化する問題は, $\|\boldsymbol{x} - P_i \boldsymbol{x}\|^2$ を最小化することと同値である. これを計算すると, $\|\boldsymbol{x} - P_i \boldsymbol{x}\|^2 = (\boldsymbol{x} - P_i \boldsymbol{x})^T (\boldsymbol{x} - P_i \boldsymbol{x}) = \|\boldsymbol{x}\|^2 - 2\boldsymbol{x}^T P_i \boldsymbol{x} + \boldsymbol{x}^T P_i^T P_i \boldsymbol{x}$ となる. $P_i = U_i U_i^T$ から $P_i^T P_i = P_i$ となるので, 結局 $\|\boldsymbol{x} - P_i \boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2 - \boldsymbol{x}^T P_i \boldsymbol{x}$ となる. このうち, \boldsymbol{x} は i に依存せず最小化とは無関係であるので, $\|\boldsymbol{x}\|^2$ を無視することができ, $-\boldsymbol{x}^T P_i \boldsymbol{x}$ を最小化すればよいことが分かる. これは, $\boldsymbol{x}^T P_i \boldsymbol{x}$ を最大化する問題となつ

ている。

9. SVM の学習時に必要となる計算量・メモリ量がデータ数およびデータの次元数とどのような関係にあるかについて述べなさい。

SVM では各データ間の内積によって定義されるマトリクスを用いた 2 次形式を含む目的関数の最適化を行う。このため、データ数 n に対してマトリクスのサイズは n^2 となる。また、この最適化問題を解くための計算量は $O(n^2)$ となる。データの次元数 d は実質的にカーネル関数を計算する際にしか問題にならず、この計算量は $O(dn^2)$ となる。したがって、次元数が大きい場合でもデータ数が多くなければ、学習を行うことができる。

10. 無限にトレーニングデータが与えられる場合の最近傍識別器の識別性能の限界について説明しなさい。

最近傍識別器では、無限にデータが与えられる場合、データの真の確率密度分布から求められる不可避免的な誤り確率である Bayes 確率よりも大きく、その 2 倍未満の誤り確率が達成できる。

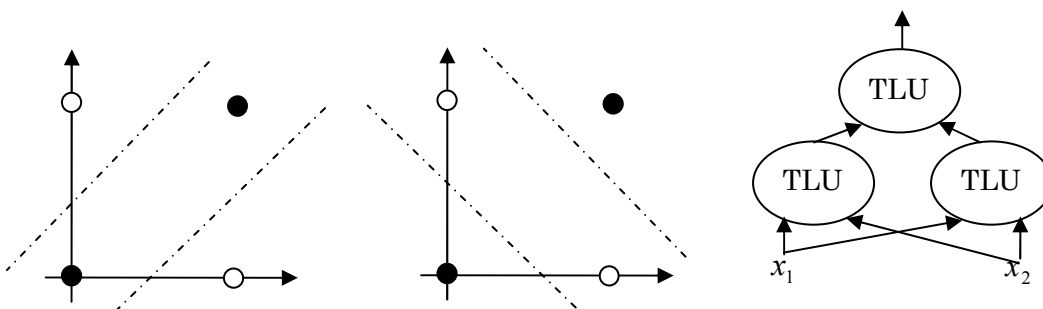
11. 最近傍探索における次元の呪縛とは何かについて述べなさい。

高次元の点分布（共分散行列のランク数が高いベクトルの分布）では、任意の点から見て他の点がほぼ等距離に位置するようになる。この結果、探索のために与えたクエリから見ても、全ての点はほぼ等距離に位置するようになり、どのような最近傍探索アルゴリズムを用いても全ての点とクエリとの間の距離を計算しなければならなくなる。これを最近傍探索における次元の呪縛と呼ぶ。

12. k-means アルゴリズムと EM アルゴリズムの類似性と相違性について述べよ。

k-means アルゴリズムと EM アルゴリズムは、1) 両者とも、各データとクラスタ間の帰属関係を暫定的に求める E ステップと、その帰属に基づいて各クラスタを規定する M ステップの計算を行っており、本質的には同種のアプローチと見なすことができる。他にも 2) 計算の前にクラスタの個数をあらかじめ決めなければならない 3) 初期値に依存して計算結果が変化する、といった共通の特性を持っている。この一方で、k-means クラスタリングでは、各データはある一つのクラスタのみに属することを前提とするが、EM では複数の分布にどの程度帰属するのかを表わす数値を各データが持っているという点が異なる。

13. Exclusive OR のパターンを学習することができる TLU ネットワークを図示し、その構造の必然性を説明しなさい。



TLU は線形識別面の学習を行う機構である。線形識別面で Exclusive OR のパターンを表現しようとする、少なくとも 2 つの TLU が必要になる。どちらのケースでも、線形識別面によって区切られるどちらの半空間にパターンが属するのかを判定する TLU と、それらを総合して、出力として 1 を出すか否かを決定する TLU が存在する必要があるため、上図の構造のネットワークになる。但し、各 TLU の -1 に固定された入力省略している。