2009/02/04　パターン認識特論試験問題　学籍番号＿＿＿＿＿＿＿＿＿＿氏名＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

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［　　］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. 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。(10 点)

3. 混合正規分布を推定する際に用いられる EM アルゴリズムとクラスタリングに用いられる k-means アルゴリズムの類似点について述べなさい．(5 点)

4. 下記の表のようなテスト-目標属性間の関係があるとき, エントロピーゲインに基づいて決定木を求めなさい．(10 点)

T1	T2	T3	T4	目標属性
1	0	1	1	1
1	1	1	1	1
1	1	1	0	0
1	0	1	0	0
1	1	0	1	0
1	0	0	1	0
1	1	0	1	0
0	1	0	1	1
0	0	1	1	1
0	1	0	1	1
0	0	0	1	1
0	0	1	0	0
0	1	0	0	0
0	0	1	0	0

5. d 次元ベクトル $\boldsymbol{f}_i$ $(i=1,\cdots,M)$ 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．$F=\left[\boldsymbol{f}_i \quad \cdots \quad \boldsymbol{f}_M\right]$ 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^TF 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．(5 点)

6. n 個のデータ $\boldsymbol{x}_i$ に対する自己相関行列を $R=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i{}^T$，平均を $\bar{\boldsymbol{x}}$ で表す．このデータに対する，共分散行列 Σ を R と $\bar{\boldsymbol{x}}$ を用いて表しなさい．(10 点)

7. パターン $\mathbf{x}$ を正規直交基底 $\mathbf{u}_i (i = 1, \cdots, n)$ が張る部分空間に正射影する行列 P を $U = \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{pmatrix}$ という行列を用いて表現しなさい．(5 点)

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（$i = 1, \cdots, C$）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．(10 点)

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．(5 点)

10. 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．(8 点)

11. 判別分析では，フィッシャー比 $\dfrac{\mathbf{A}^T \Sigma_B \mathbf{A}}{\mathbf{A}^T \Sigma_W \mathbf{A}}$ を最大化する問題を解き，射影軸 $\mathbf{A}$ を求める． この問題の解き方を示しなさい．(12 点)

2009/02/04　パターン認識特論演習問題　学籍番号　解答例　氏名＿＿＿＿＿＿＿＿＿＿＿＿

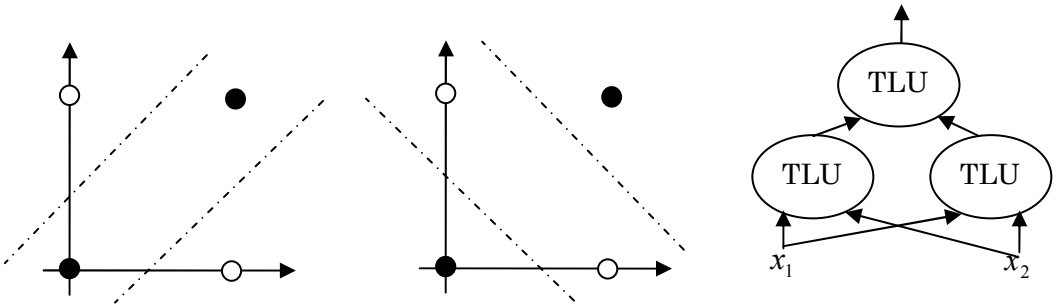
1. 下記の文章の空欄を埋めなさい．（小さい四角は 1 問 1 点，大きい四角は 3 点，合計 20 点）

- 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．

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

- 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［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，［ラグランジェの未定係数法］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，［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. 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。(10 点)



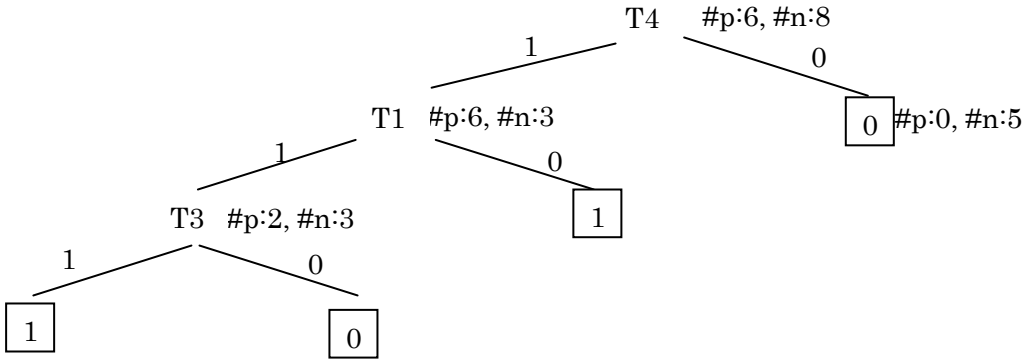
TLU は線形識別面の学習を行う機構である。線形識別面で Exclusive OR のパターンを表現しようとすると、少なくとも 2 つの TLU が必要になる。どちらのケースでも、線形識別面によって区切られるどちらの半空間にパターンが属するのかを判定する TLU と、それらを総合して、出力として 1 を出すか否かを決定する TLU が存在する必要があるため、上図の構造のネットワークになる。但し、各 TLU の-1 に固定された入力は省略している。

3. 混合正規分布を推定する際に用いられる EM アルゴリズムとクラスタリングに用いられる k-means アルゴリズムの類似点について述べなさい．(5 点)

k-means アルゴリズムは，1）各データを最も近い位置にあるクラスタ中心に対応付け，グループ化する．この 2）グループ毎に重心の位置を求め，それを新たなクラスタ中心とする，という計算を繰り返すクラスタリング手法である．一方，EM アルゴリズムは，1）各データに対し，各分布への帰属度を求め，2）この帰属度のもとで各分布のパラメータの最尤推定を行う，という計算を反復する．これらの計算は，どちらも 1)の段階で，データとクラスタ（または分布）の対応関係を暫定的に決め，その対応関係の下で 2）クラスタ（または分布）を規定するパラメータ（クラスタ中心，平均・分散など）を求めるという枠組みでとらえられる．

4. 下記の表のようなテスト-目標属性間の関係があるとき，エントロピーゲインに基づいて決定木を求めなさい．(10 点)

T1	T2	T3	T4	目標属性
1	0	1	1	1
1	1	1	1	1
1	1	1	0	0
1	0	1	0	0
1	1	0	1	0
1	0	0	1	0
1	1	0	1	0
0	1	0	1	1
0	0	1	1	1
0	1	0	1	1
0	0	0	1	1
0	0	1	0	0
0	1	0	0	0
0	0	1	0	0



5. d 次元ベクトル $\boldsymbol{f}_i\ (i=1,\cdots,M)$ 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．$F=\left[\boldsymbol{f}_i \quad \cdots \quad \boldsymbol{f}_M\right]$ 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^TF 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．(5 点)

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

6. n 個のデータ $\boldsymbol{x}_i$ に対する自己相関行列を $R=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^{\,T}$，平均を $\bar{\boldsymbol{x}}$ で表す．このデータに対する，共分散行列 Σ を R と $\bar{\boldsymbol{x}}$ を用いて表しなさい．(10 点)

$\Sigma=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})^T=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i\boldsymbol{x}_i^{\,T}-\bar{\boldsymbol{x}}\boldsymbol{x}_i^{\,T}-\boldsymbol{x}_i\bar{\boldsymbol{x}}^T+\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T)=R-2\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T+\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T=R-\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T$ となる．

7. パターン $\mathbf{x}$ を正規直交基底 $\mathbf{u}_i (i=1,\cdots,n)$ が張る部分空間に正射影する行列 P を $U=\begin{pmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{pmatrix}$ という行列を用いて表現しなさい．(5 点)

$$P=UU^T$$

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（$i=1,\cdots,C$）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．(10 点)

$\| \boldsymbol{x}-P_i\boldsymbol{x} \|$ 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，$\| \boldsymbol{x}-P_i\boldsymbol{x} \|^2=(x-P_ix)^T(x-P_ix)=\| x \|^2-2x^TP_ix+x^TP_i^TP_ix$ となる．$P_i=U_iU_i^T$ から $P_i^TP_i=P_i$ となるので，結局 $\| \boldsymbol{x}-P_i\boldsymbol{x} \|^2=\| x \|^2-x^TP_ix$ となる．このうち，x は i に依存せず最小化とは無関係であるので，$\| x \|^2$ を無視することができ，$-x^TP_ix$ を最小化すればよいことが分かる．これは，x^TP_ix を最大化する問題となっている．

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．(5 点)

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. 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．(8 点)

$$\Sigma = V\Lambda V^T = \begin{bmatrix}\varphi_1 & \cdots & \varphi_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \lambda_n\end{bmatrix}\begin{bmatrix}\varphi_1^T\\ \vdots\\ \varphi_n^T\end{bmatrix} = \sum_{i=1}^{n}\lambda_i\varphi_i\varphi_i^T$$

11. 判別分析では，$\Sigma_W=\dfrac{n_1\Sigma_1+n_2\Sigma_2}{n_1+n_2}$，$\Sigma_B=\dfrac{n_1n_2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T}{(n_1+n_2)^2}$ とするとき,フィッシャー比 $\dfrac{\mathbf{A}^T\Sigma_B\mathbf{A}}{\mathbf{A}^T\Sigma_W\mathbf{A}}$ を最大化する問題を解き，射影軸 $\mathbf{A}$ を求める． この問題の解き方を示しなさい．(12 点)

$\mathbf{A}^T\Sigma_W\mathbf{A}=1$ という条件の下での $\mathbf{A}^T\Sigma_B\mathbf{A}$ の最大化問題と見なすと，Lagrange の未定係数法により、次の目的関数が得られる．$J(\mathbf{A})=\mathbf{A}^T\Sigma_B\mathbf{A}-\lambda(\mathbf{A}^T\Sigma_W\mathbf{A}-1)$ この式の両辺を $\mathbf{A}$ で微分した式，$\dfrac{\partial}{\partial\mathbf{A}}J(\mathbf{A})=2\Sigma_B\mathbf{A}-2\lambda\Sigma_W\mathbf{A}$ は 0 となることから，$\Sigma_B\mathbf{A}=\lambda\Sigma_W\mathbf{A}$ となり，これを整理することにより，$(\Sigma_W^{-1}\Sigma_B-\lambda I)\mathbf{A}=0$ が得られる．即ち，この問題は，$\Sigma_W^{-1}\Sigma_B$ の固有値問題に帰着する．具体的には，$\mathbf{A}^T\Sigma_B\mathbf{A}=\lambda\mathbf{A}^T\Sigma_W\mathbf{A}=\lambda$ であることから，$\Sigma_W^{-1}\Sigma_B$ の最大固有値が $J(\mathbf{A})$ の最大値となり，それに対応する固有ベクトルが $\mathbf{A}$ となる．