2014/01/31　パターン認識特論演習問題　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. d 次元空間中に n 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 と n を用いて表わしなさい．（5 点）

3. 共分散行列 $\boldsymbol{\Sigma}$ 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，$\boldsymbol{\varphi}_i^T \boldsymbol{\Sigma} \boldsymbol{\varphi}_i$ の値はいくらになるか計算で求めなさい．（5 点）

4.　２クラスの判別分析では，クラス内分散 Σ_W とクラス間分散 Σ_B を用いて定義されるフィッシャー比 $\boldsymbol{F}$ を最大化する問題を解くことで特徴ベクトルに対する射影軸を求める．　クラス１，２，のデータ数を n_1, n_2 ,共分散行列を Σ_1, Σ_2 ,平均値を $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$，射影軸を $\boldsymbol{A}$ としたとき，次の問いに答えなさい．（５＋５＝１０点）

（１） $\Sigma_W, \Sigma_B, \boldsymbol{F}$ をこれらの記号を用いて表しなさい．

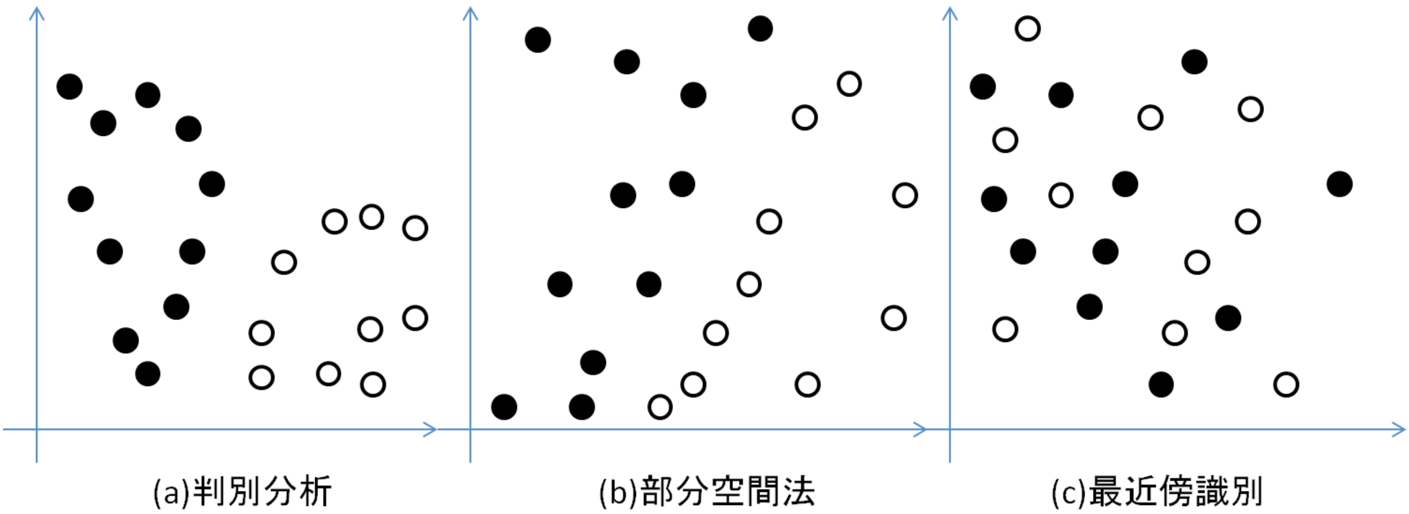
（２） $\boldsymbol{F}$ を最大化する $\boldsymbol{A}$ の求め方を示しなさい．

5.　$f(x)$ の Fourier 変換を $F(\omega)=\int_{-\infty}^{\infty} f(x)e^{-jwx}dx$ とする．このとき，次の問いに答えなさい．（５＋５＝１０点）

（１）$f(x)$ が平行移動したパターン $g(x)=f(x+\mu)$ の Fourier 変換 $G(\omega)$ を $F(\omega)$ を用いた式で表しなさい．

（２）この結果からパワーや自己相関を用いずに平行移動に対する不変量を求める手順の概略を述べなさい．

6.　下記の 2 次元特徴分布それぞれに対して用意された識別方法の中で不適切なものを選び，適切な識別器と，その理由を述べよ．　（10 点）



理由：

7.　SVM や ADA Boosting が適用しやすい問題と，しにくい問題の例を挙げ，その理由を説明しなさい．（５点）

8.　ホワイトニングの計算と Mahalanobis 距離の計算の関係を式を用いて説明しなさい．（１５点）

9.　直交展開における各基底関数の係数を内積によって決定することが，最小二乗法として妥当である事を示しなさい．（２０点）

2013/02/06　パターン認識特論試験問題解答例

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 $\boxed{\text{0－1 損失}}$ 基準のもとで $\boxed{\text{損失}}$ 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 $\boxed{\text{Bayes}}$ 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，$\boxed{\text{確率密度関数 } p(x|\omega_i)}$ 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．

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

- 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 $\boxed{\text{余弦}}$ 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 $\boxed{\text{複合類似度法}}$ では，$\boxed{\text{自己相関行列}}$ 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 $\boxed{\text{部分空間法}}$ 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 $\boxed{\text{角度}}$ に基づいて識別を行っているからである．
- これ以外に，$\boxed{\text{最近傍識別器}}$ 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，$\boxed{\text{Voronoi Condensing}}$ 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 $\boxed{\text{SVM}}$ 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，$\boxed{\text{ラグランジェの未定係数法}}$ 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 $\boxed{\alpha \neq 0}$ となるパターンは特殊例すなわち $\boxed{\text{サポートベクター}}$ となる．
- この例以外にも $\boxed{\text{ADA Boosting}}$ 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，$\boxed{\text{最近傍識別}}$ 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. d 次元空間中に n 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 と n を用いて表わしなさい．（5 点）

$$\min(d, n-1)$$

3. 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，$\boldsymbol{\varphi}_i^T \boldsymbol{\Sigma} \boldsymbol{\varphi}_i$ の値はいくらになるか計算で求めなさい．（５点）

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

4.　２クラスの判別分析では，クラス内分散 Σ_W とクラス間分散 Σ_B を用いて定義されるフィッシャー比 F を最大化する問題を解くことで特徴ベクトルに対する射影軸を求める．　クラス１，２，のデータ数を n_1, n_2,共分散行列を Σ_1, Σ_2,平均値を $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$，射影軸を $\boldsymbol{A}$ としたとき，次の問いに答えなさい．（５＋５＝１０点）

（１） Σ_W, Σ_B, F をこれらの記号を用いて表しなさい．

$$\Sigma_W = \frac{n_1\Sigma_1 + n_2\Sigma_2}{n_1 + n_2}$$

$$\Sigma_B = \frac{n_1 n_2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T}{(n_1 + n_2)^2}$$

$$F = \frac{\mathbf{A}^T\Sigma_B\mathbf{A}}{\mathbf{A}^T\Sigma_W\mathbf{A}}$$

（２） F を最大化する $\boldsymbol{A}$ の求め方を示しなさい．

$\mathbf{A}^T\Sigma_W\mathbf{A} = 1$ 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，$\frac{\partial}{\partial \mathbf{A}}J(\mathbf{A}) = 2\Sigma_B\mathbf{A} - 2\lambda\Sigma_W\mathbf{A}$ 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．

5.　$f(x)$ の Fourier 変換を $F(\omega)=\int_{-\infty}^{\infty} f(x)e^{-jwx}dx$ とする．このとき，次の問いに答えなさい．（５＋５＝１０点）

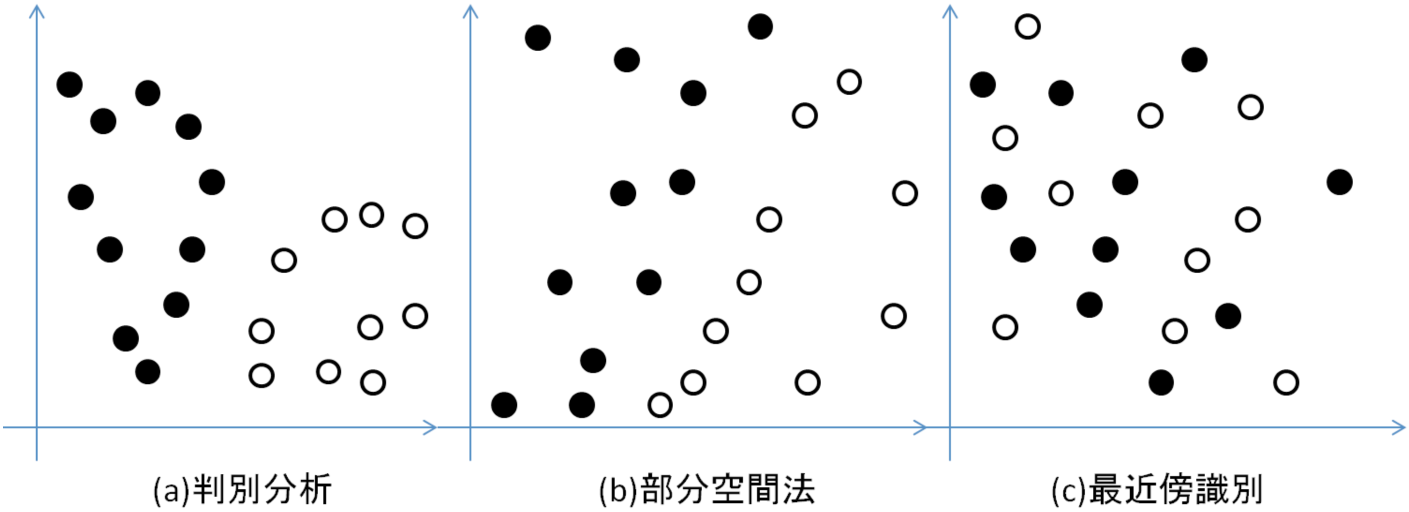
（１）$f(x)$ が平行移動したパターン $g(x) = f(x+\mu)$ の Fourier 変換 $G(\omega)$ を $F(\omega)$ を用いた式で表しなさい．

$$G(\omega) \equiv \int_{-\infty}^{\infty} f(x+\mu)e^{-jwx}\,dx = \int_{-\infty}^{\infty} f(X)e^{-jw(X-\mu)}\,dX = e^{jw\mu}\int_{-\infty}^{\infty} f(X)e^{-jwX}\,dX = e^{jw\mu}F(\omega)$$

（２）この結果からパワーや自己相関を用いずに平行移動に対する不変量を求める手順の概略を述べなさい．

上式の通り，平行移動は回転因子 $e^{jw\mu}$ を $F(\omega)$ に乗ずる形で求められる．この対数をもとめると，$\log_e e^{jw\mu}F(\omega) = j\omega\mu + \log_e F(\omega)$ となる．この式から，平行移動は $j\omega\mu$ の項に現れる．
したがって， $j\omega\mu$ の成分を $G(\omega)$ から除去する，つまり， $j\omega\mu$ の直交補空間に $G(\omega)$ を射影すれば良い．

6.　下記の 2 次元特徴分布それぞれに対して用意された識別方法の中で不適切なものを選び，適切な識別器と，その理由を述べよ．　（10 点）



理由：(b) の特徴分布にそのまま部分空間法を適用するのは適切ではない．これは，識別境界が原点を通過しないため，自己相関行列の固有値分解で得られる基底群の張る部分空間への射影をしても，原点付近のデータは正しき識別できないためである．判別分析や SVM，最近傍識別，正規分布の当てはめによる統計的識別などを用いるべきである．

7.　SVM や ADA Boosting が適用しやすい問題と，しにくい問題の例を挙げ，その理由を説明しなさい．（５点）

SVM や ADA Boosting は基本的に 2 クラスの識別問題に適用される判別モデルである．したがって，クラス数 N が極めて多い問題では，各クラス間の隣接関係を調べる必要がある．このチェックのためには ${}_N C_2$ 組の組み合わせを調べなければならない．さらに，識別時には決定木を構築し，どの順序で識別を行うべきかを考えなければならない．したがって，N が数万のオーダーになる漢字の文字認識などには適さないと言える．

8.　ホワイトニングの計算と Mahalanobis 距離の計算の関係を式を用いて説明しなさい．（１５点）

N 個のデータ，$\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ から平均値 $\boldsymbol{\mu}$ を求め，共分散行列 Σ を求める．これらを用いて計算される平均 $\boldsymbol{\mu}$ から $\boldsymbol{x}$ までの距離 $(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ が，Mahalanobis 距離である．これは，平均 $\boldsymbol{\mu}$，共分散行列 Σ を母数とする正規分布の等確率密度面で同じ値を取り，$\boldsymbol{x}$ が $\boldsymbol{\mu}$ から離れるほど大きな値になるという性質を持つ．
これに対して，ホワイトニングとはデータの分布を変換する手法であり，Mahalanobis 距離の式から以下のように導出される．

$$(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=(\boldsymbol{x}-\boldsymbol{\mu})^T V\Lambda^{-1}V^T(\boldsymbol{x}-\boldsymbol{\mu})=\left(\Lambda^{-\frac{1}{2}}V^T(\boldsymbol{x}-\boldsymbol{\mu})\right)^T\left(\Lambda^{-\frac{1}{2}}V^T(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

この $\Lambda^{-\frac{1}{2}}V^T(\boldsymbol{x}-\boldsymbol{\mu})$ が，ホワイトニングである．但し，Λ は Σ の固有値から成る対角行列，V は Σ の固有ベクトルから成る直交行列 $V=\begin{pmatrix}\boldsymbol{\varphi}_1 & \cdots & \boldsymbol{\varphi}_N\end{pmatrix}$，$\Lambda^{-\frac{1}{2}}$ は Σ の固有値の平方根（つまり偏差）の逆数から成る対角行列である．この式から言える事は，ホワイトニングでは，平均値を引いたベクトルを回転させて各ベクトルの成分間の相関を消し，それぞれの偏差で除する事によって，無相関かつ等方性の分布を得ることができるということである．

9.　直交展開における各基底関数の係数を内積によって決定することが，最小二乗法として妥当である事を示しなさい．（２０点）

未知の関数を $f(x)$ とし， $f^*(x)=\sum_{i=1}^{N} c_i\varphi_i(x)$ という形式でこの関数を近似する．このとき，

$\int_a^b \left(f(x)-f^*(x)\right)^2 dx$ の値を最小化する　c_i を求める問題が最小二乗法である．

$(f,g)=\int_a^b f(x)g(x)dx$ を $f(x)$ と $g(x)$ の内積と定義する．

このとき，最小化する目的値は

$$\int_a^b \left(f(x)-f^*(x)\right)^2 dx = \| f(x)-f^*(x) \|^2$$

$$= \| f(x)-\sum_{i=1}^{N} c_i\varphi_i(x) \|^2$$

$$= \| f(x) \|^2 -2(f(x),\sum_{i=1}^{N} c_i\varphi_i(x)) + \| \sum_{i=1}^{N} c_i\varphi_i(x) \|^2$$

となる．これを c_i で偏微分して０とおく事により，右の結果を得る．これを連立方程式で表せば，下記のようになり

$$\frac{\partial}{\partial c_k} \| f(x)-\sum_{i=1}^{N} c_i\varphi_i(x) \|^2$$

$$= \frac{\partial}{\partial c_k} \| f(x) \|^2 -2(f(x),\sum_{i=1}^{N} c_i\varphi_i(x)) + \| \sum_{i=1}^{N} c_i\varphi_i(x) \|^2$$

$$= -2(f(x),\varphi_k(x)) + 2(\varphi_k(x),\sum_{i=1}^{N} c_i\varphi_i(x)) = 0$$

$$\begin{bmatrix} (\varphi_1,\varphi_1) & \cdots & (\varphi_1,\varphi_N) \\ \vdots & \ddots & \vdots \\ (\varphi_N,\varphi_1) & \cdots & (\varphi_N,\varphi_N) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} (\varphi_1,f) \\ \vdots \\ (\varphi_N,f) \end{bmatrix}$$

直交関数を用いた場合，左辺の行列は単位行列になり，結局，下記のようになる．

$$\begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} (\varphi_1,f) \\ \vdots \\ (\varphi_N,f) \end{bmatrix}$$

したがって，最小二乗法で求めた最適な係数ベクトルは直交関数系を用いた場合，対象とする関数との単純な内積によって表される．