2014/02/05　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［　　　　］基準のもとで損失を［　　］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，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．

  ［　　　　　　　　　　　　　　　　　　　　］

- 一方，［　　　］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内積を計算し，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［　　　］類似度法では，［　　　　］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．さらに，「犬」「大」「太」のように，大局的には似ていても僅かなパターンの違いで所属するクラスが変化するようなケースに対処するため，他の類似クラスとの類似性をペナルティとする［　　］類似度法もある．
- 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．このほかにも，［　　　　］グラフや，RNG を用いたプロトタイプの削減法がある．
- 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自己相関行列を求めたとき，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 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，$\boldsymbol{\varphi}_i^T \Sigma^{-1} \boldsymbol{\varphi}_i$ の値はいくらになるか計算で求めなさい．（5 点）

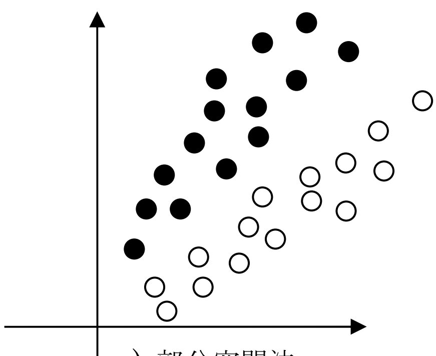
4. d 次元ベクトル $\boldsymbol{f}_i \ (i=1,\cdots,M)$ が与えられている．その固有ベクトルを求める方法について考える．$F=\left[\boldsymbol{f}_i \quad \cdots \quad \boldsymbol{f}_M\right]$ とすると，共分散行列を M 倍した行列は $\sum_{i=1}^{M} \boldsymbol{f}_i \boldsymbol{f}_i^T = FF^T$ と表せる．この行列の固有ベクトルを FF^T ではなく $F^T F$ の固有ベクトルから求める方法と，その利点について述べなさい．（10 点）

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）

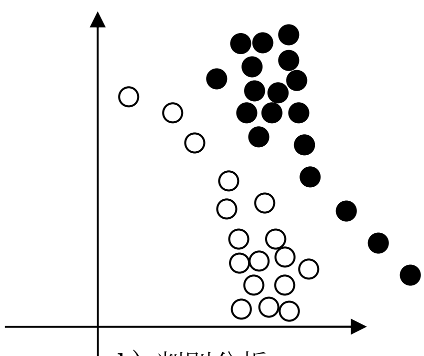
（１）$f(x)$ が平行移動したパターン $g(x)=f(x-\mu)$ の Fourier 変換 $G(\omega)$ を $F(\omega)$ を用いた式で表しなさい．

（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．

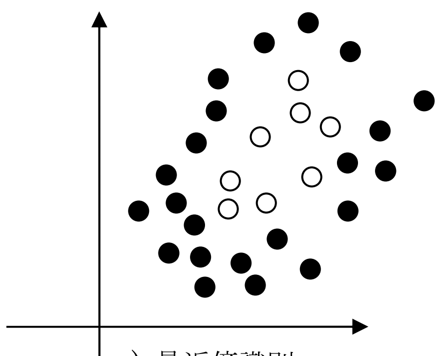
6. 下記の 2 次元特徴分布それぞれに対して用意された識別方法の中で不適切なものを選び，適切な識別器と，その理由を述べよ．（10 点）



a）部分空間法



b）判別分析



c）最近傍識別

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 汎距離の計算の関係を式を用いて説明しなさい．（１５点）

9.　カーネル主成分分析は，通常の主成分分析で非正則な共分散行列の固有値問題を解かなければならなくなるケースでも，この問題を回避できるという利点がある．その理由について説明しなさい．(5 点)

10. 二入力の Exclusive OR の挙動を単一の TLU で模倣することはできないことを示した上で， TLU を階層的に結合したネットワークによって Exclusive OR の挙動が学習できることを，図を用いて示しなさい．また，この図における各 TLU の働きも説明しなさい． （7 点）

11. パターン $\boldsymbol{x}$ 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．また，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．（4+4=8 点）

2014/02/05　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 $\boxed{\text{0－1 損失}}$基準のもとで損失を$\boxed{\text{最小}}$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，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，$\boxed{\text{単純}}$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内積を計算し，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．さらに，「犬」「大」「太」のように，大局的には似ていても僅かなパターンの違いで所属するクラスが変化するようなケースに対処するため，他の類似クラスとの類似性をペナルティとする$\boxed{\text{混合}}$類似度法もある．
- 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．このほかにも，$\boxed{\text{Gabriel}}$グラフや，RNG を用いたプロトタイプの削減法がある．
- 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自己相関行列を求めたとき，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）

$$\min(d, n)$$

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，$\boldsymbol{\varphi}_i^T \Sigma^{-1} \boldsymbol{\varphi}_i$ の値はいくらになるか計算で求めなさい．（5 点）

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

4. 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．(10 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，$d << M$ 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．

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)

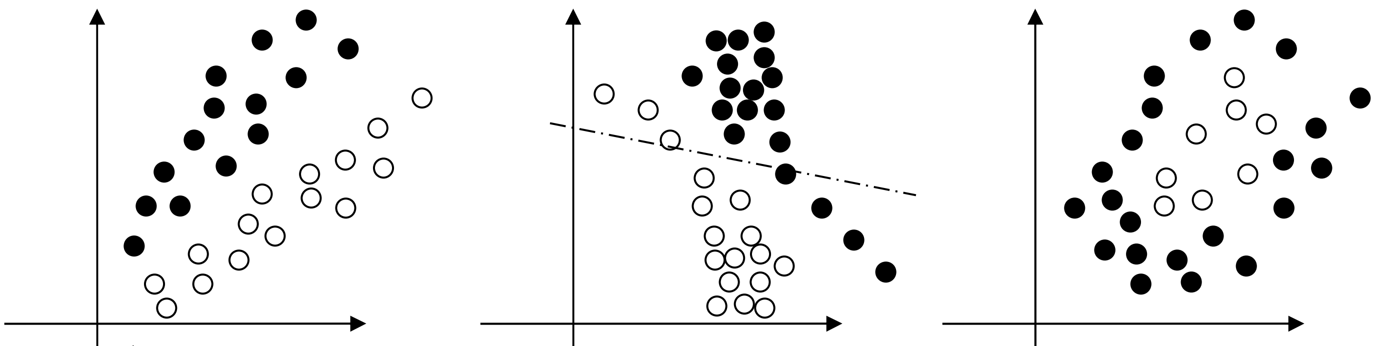
（１）$f(x)$ が平行移動したパターン $g(x)=f(x-\mu)$ の Fourier 変換 $G(\omega)$ を $F(\omega)$ を用いた式で表しなさい．

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

（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)



a) 部分空間法　　b) 判別分析　　c) 最近傍識別

理由：b) 判別分析は不適切．その理由は各データの分布が集中しているため，図のように水平に近い識別面が得られるにもかかわらず，実際の識別面はかなり傾斜しているためである．このデータ分布に対しては Linear SVM を用いるのが適している．

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 汎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 $\boldsymbol{x}$ までの間で定義される尺度

$$(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})$$

が，Mahalanobis 汎距離である．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 汎距離の式から以下のように導出される．

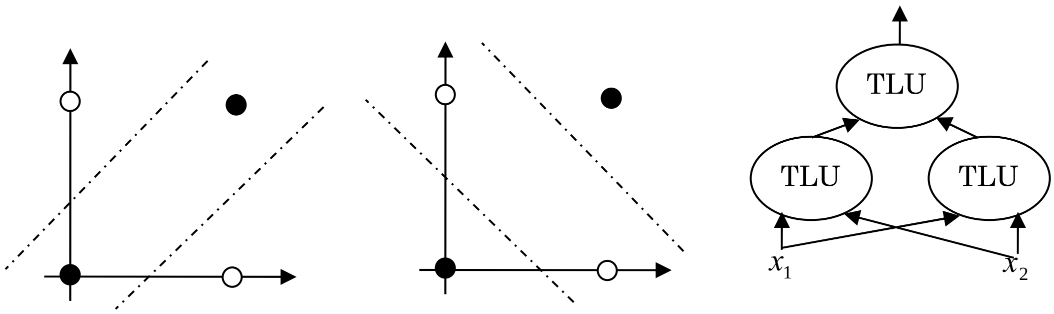
$$(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}-\boldsymbol{\mu})^T V\Lambda^{-1}V^T (\boldsymbol{x}-\boldsymbol{\mu}) = \left(\Lambda^{-\frac{1}{2}}V^T(\boldsymbol{x}-\boldsymbol{\mu})\right)^T \left(\Lambda^{-\frac{1}{2}}V^T(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

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. カーネル主成分分析は，通常の主成分分析で非正則な共分散行列の固有値問題を解かなければならなくなるケースでも，この問題を回避できるという利点がある．その理由について説明しなさい．（5 点）

カーネル主成分分析では，主軸はトレーニングデータの線形結合で表現されることを仮定し，トレーニングデータ間のカーネル関数を要素とする「グラム行列」の固有値問題を解くことで，この線形結合係数を求めるという計算が行われる．このため，次元数＋１よりも少ないサンプルしか与えられない場合でも，グラム行列が正則である限り，安定に固有値問題を解くことができる．

10. 二入力の Exclusive OR の挙動を単一の TLU で模倣することはできないことを示した上で， TLU を階層的に結合したネットワークによって Exclusive OR の挙動が学習できることを，図を用いて示しなさい．また，この図における各 TLU の働きも説明しなさい．（7 点）



(a) (b) (c)

上図(a)(b)のように，二入力 Exclusive OR は(0,0)と(1,1)で 0(●で表示)，(0,1)と(1,0)で 1（○で表示）となる．TLU は線形識別面の学習を行う機構であるので，0 と 1 の間に境界となる線形識別面を入れると，上図一点鎖線で表示したように，一つの線形識別面で境界を表現することはできない．少なくとも 2 つの TLU が必要になる。上図(a)(b)の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 が二つ必要であり，それら 2 つの TLU の出力を総合して、出力として 1 を出すか否かを決定する TLU が最終的な出力層に必要となる．但し上図(c)では各 TLU の-1 に固定された入力は省略している．

11. パターン $\boldsymbol{x}$ 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．また，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．（4+4=8 点）

$$P = UU^T$$

$$\|\boldsymbol{x}-P_i\boldsymbol{x}\|^2 = (\boldsymbol{x}-P_i\boldsymbol{x})^T(\boldsymbol{x}-P_i\boldsymbol{x}) = \|\boldsymbol{x}\|^2 - 2\boldsymbol{x}^T P_i \boldsymbol{x} + \boldsymbol{x}^T P_i^T P_i \boldsymbol{x}$$

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．したがって，$\|\boldsymbol{x}-P_i\boldsymbol{x}\|^2$ を最小化する問題は， $\boldsymbol{x}^T P_i \boldsymbol{x}$ を最大化する問題となっている．