2012/02/08　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．(13 点)

- 古典的パターン認識理論では，入力 $\boldsymbol{x}$ 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 $\boxed{\qquad\qquad}$ を用いる場合はこの値の大小で識別を行うが，これは $\boxed{\qquad\qquad}$ 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．
- 単純類似度法では，入力と各クラスを代表するパターンとの $\boxed{\qquad}$ 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，$\boxed{\qquad\qquad}$ 行列の固有ベクトルから成る複数の代表パターンとの $\boxed{\qquad\qquad}$ 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．この方法では，「犬」「大」「太」など類似パターン間での誤識別が起きるため，誤識別を起こしやすいパターンに対するペナルティを導入した $\boxed{\qquad\qquad}$ も用いられる．
- これ以外に，最近傍識別器では，入力パターンを，それに最も近い既知のパターン（プロトタイプ）が属するクラスに分類する．この方法では，$\boxed{\qquad\qquad\qquad}$ という手法を用いると，識別面を変えずに記憶するプロトタイプを減らすことができる．同様の方法として，$\boxed{\qquad\qquad\qquad}$, RNG Editing がある．これらの処理の結果，識別境界付近の誤識別を起こしやすい特殊例が残される．
- 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 $\boxed{\qquad}$ 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 $\boxed{\qquad}$ となるパターンは特殊例すなわち $\boxed{\qquad\qquad}$ となる．
- この例以外にも $\boxed{\qquad\qquad}$ 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，$\boxed{\qquad\qquad}$ 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.　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

3.　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．(7 点)

4. 共分散行列$\mathbf{\Sigma}$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，$\mathbf{\Sigma}$をλ_iと$\boldsymbol{\varphi}_i$で，表しなさい．（10 点）

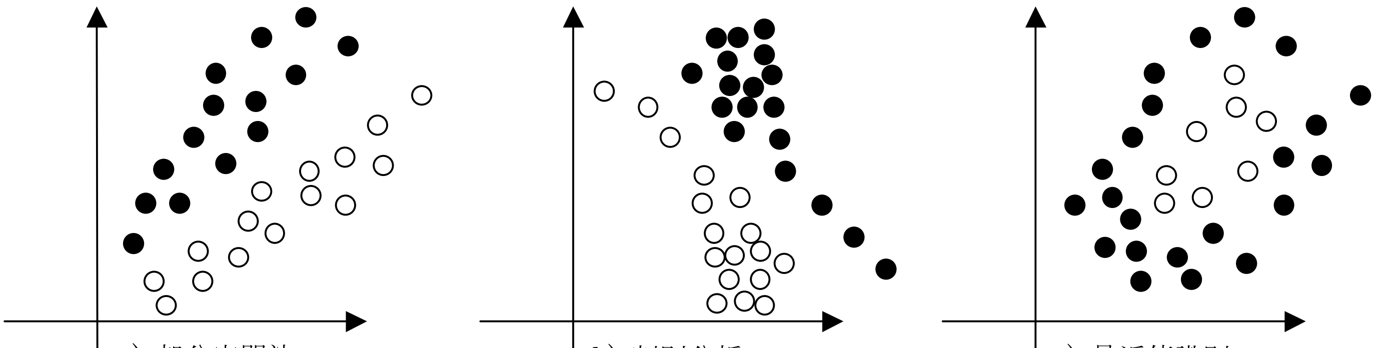
5. 判別分析では，$\mathbf{\Sigma}_W = \dfrac{n_1\mathbf{\Sigma}_1 + n_2\mathbf{\Sigma}_2}{n_1+n_2}$，$\mathbf{\Sigma}_B = \dfrac{n_1 n_2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T}{(n_1+n_2)^2}$とするとき,フィッシャー比$\dfrac{\mathbf{A}^T\mathbf{\Sigma}_B\mathbf{A}}{\mathbf{A}^T\mathbf{\Sigma}_W\mathbf{A}}$を最大化する問題を解き，射影軸$\mathbf{A}$を求める． 2 クラスの識別問題における，この問題の解き方を示しなさい．（10 点）

6. $f(x)$の Fourier 変換を$F(\omega)=\int_{-\infty}^{\infty} f(x)e^{-jwx}dx$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，$\mu$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．（10 点）

7. 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．$F=\begin{bmatrix} \boldsymbol{f}_i & \cdots & \boldsymbol{f}_M \end{bmatrix}$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）

8.　カーネル SVM を用いた識別では，識別面の形状が複雑になるほど計算時間がかかる．この理由を説明せよ．（5 点）

9.　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．（5 点）



a）部分空間法　b）判別分析　c）最近傍識別

理由：

10. 漢字の認識に SVM や ADA Boosting が適用しにくい理由にについて述べよ．（3 点）

11. 不偏推定と最尤推定の意味の違いについて述べ，それぞれによって分散と平均を求めるとどのようになるか式で示せ．（5 点）

12. 区間 $[\Delta, M\Delta]$ で定義された関数 $f(x)$ を $f^*(x) = \sum_{i=1}^{N} c_i \varphi_i(x)$ で近似する問題を考える． $x_k = k\Delta$ ，($k = 1, \cdots, M$)における $f(x)$ の観測値が得られる場合，最小二乗法に基づく係数 c_i の決定法を説明しなさい．また，$\sum_{k=1}^{M} \varphi_i(x_k)\varphi_j(x_k) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ という条件が満たされる場合の係数の決定法についても説明しなさい．

（12 点）

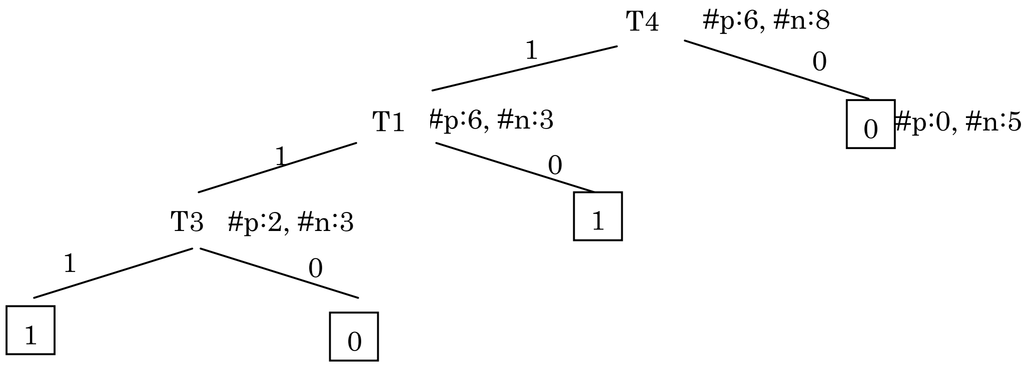
2010/02/08　パターン認識特論テスト　学籍番号＿＿解答例＿＿氏名＿＿＿＿＿＿＿＿＿＿＿＿

1.　下記の文章の空欄を埋めなさい．（13 点）

- 古典的パターン認識理論では，入力 $\boldsymbol{x}$ 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 $\boxed{P(\omega_i \mid \boldsymbol{x})}$ 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．
- 単純類似度法では，入力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．　この方法では，「犬」「大」「太」など類似パターン間での誤識別が起きるため，誤識別を起こしやすいパターンに対するペナルティを導入した 混合類似度法 も用いられる．
- これ以外に，最近傍識別器では，入力パターンを，それに最も近い既知のパターン（プロトタイプ）が属するクラスに分類する．この方法では， Voronoi Condensing という手法を用いると，識別面を変えずに記憶するプロトタイプを減らすことができる．同様の方法として， Gabriel Editing, RNG Editing があるが，これらの処理の結果，識別境界付近の誤識別を起こしやすい特殊例が残される．
- 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 $\neq 0$ 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．

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



3.　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．（7 点）

$\min(d, n-1)$

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$\boldsymbol{\varphi}_i$ $(i = 1 \cdots n)$であるとき，Σをλ_iと$\boldsymbol{\varphi}_i$で，表しなさい．(10 点)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \varphi_1 & \cdots & \varphi_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} \varphi_1^T \\ \vdots \\ \varphi_n^T \end{bmatrix} = \sum_{i=i}^{n} \boldsymbol{\lambda}_i \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^T$$

5. 判別分析では，$\Sigma_W = \dfrac{n_1\Sigma_1 + n_2\Sigma_2}{n_1 + n_2}$，$\Sigma_B = \dfrac{n_1 n_2 (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T}{(n_1 + n_2)^2}$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． 2 クラスの識別問題における，この問題の解き方を示しなさい．(10 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．

6. $f(x)$の Fourier 変換を$F(\omega) = \int_{-\infty}^{\infty} f(x)e^{-jwx}dx$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，$\mu$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．(10 点)

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

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

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．

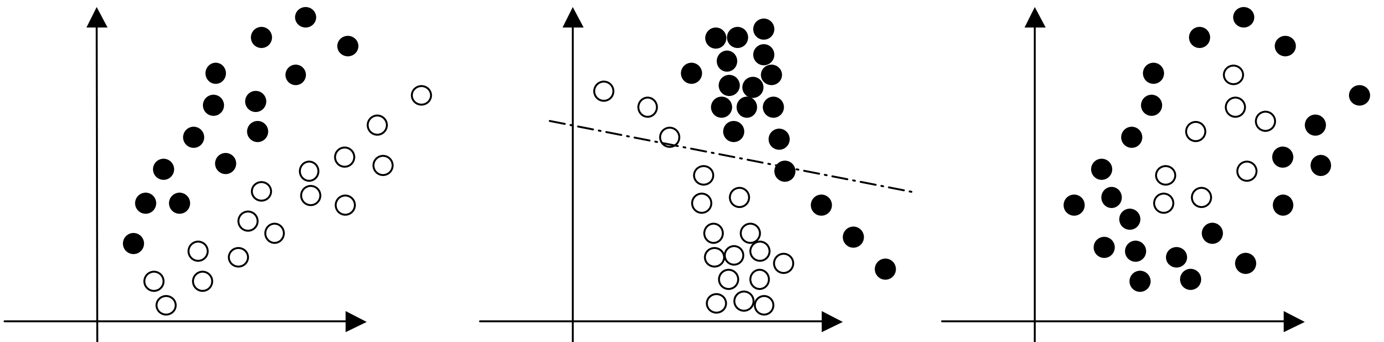
7. d次元ベクトル$\boldsymbol{f}_i$ $(i = 1, \cdots, M)$が与えられている．その固有ベクトルを求める方法について考える．$F = \begin{bmatrix} \boldsymbol{f}_i & \cdots & \boldsymbol{f}_M \end{bmatrix}$とすると，共分散行列を$M$倍した行列は$\sum_{i=1}^{M} \boldsymbol{f}_i \boldsymbol{f}_i^T = FF^T$と表せる．この行列の固有ベクトルを$FF^T$ではなく$F^TF$の固有ベクトルから求める方法と，その利点について述べなさ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

8. カーネル SVM を用いた識別では，識別面の形状が複雑になるほど計算時間がかかる．この理由を説明せよ．（5 点）

カーネル SVM では，識別の際に入力データとサポートベクターの内積（カーネル関数）を計算しなければならない．識別面の形状が複雑になればなるほど，サポートベクターの個数が増加するため，入力データとの内積（カーネル関数）の計算回数が増加するため，識別に要する時間が増加する．

9. 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．（5 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．

10. 漢字の認識に SVM や ADA Boosting が適用しにくい理由ににについて述べよ．（3 点）

漢字は数万以上のカテゴリがあり，クラス間の組み合わせで識別境界を学習する SVM や ADA Boosting は，組み合わせの数が膨大になるため，学習が困難になる．

11. 不偏推定と最尤推定の意味の違いについて述べ，それぞれによって分散と平均を求めるとどのようになるか式で示せ．（5 点）

最尤推定とは，ある分布モデルを仮定し，与えられたデータがその分布モデルから同時に生起する確率（尤度）を最大化するように分布の母数を決定する手法である．ちなみに，分散と平均を推定する場合には正規分布モデルが標準的に用いられる．これに対して，不偏推定とは，本来のデータ分布からサンプルした標本点から推定した母数の期待値がバイアスを持たず真の母数と一致するように推定する方法である．

データ数を N，各データを x_i とすると，平均は不偏推定，最尤推定ともに $\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i$ となる．これに対して，最尤推定で求めた分散は $\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2$ であるが，不偏推定では $\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2$ となる．

12. 区間 $[\Delta, M\Delta]$ で定義された関数 $f(x)$ を $f^*(x)=\sum_{i=1}^{N}c_i\varphi_i(x)$ で近似する問題を考える． $x_k=k\Delta$，($k=1,\cdots,M$)における $f(x)$ の観測値が得られる場合，最小二乗法に基づく係数 c_i の決定法を説明しなさい．また，$\sum_{k=1}^{M}\varphi_i(x_k)\varphi_j(x_k)=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$ という条件が満たされる場合の係数の決定法についても説明しなさい．（12 点）

誤差関数を$E(c_1,\cdots,c_N)=\sum_{k=1}^{M}\left(f(x_k)-\sum_{i=1}^{N}c_i\varphi_i(x_k)\right)^2$としたとき，$\frac{\partial}{\partial c_l}E(c_1,\cdots,c_N)=0$を，$l=1,\cdots,N$について成り立たせる$c_1,\cdots,c_N$を求めれば良い．

$$E(c_1,\cdots,c_N)=\sum_{k=1}^{M}\left(f(x_k)-\sum_{i=1}^{N}c_i\varphi_i(x_k)\right)^2=\sum_{k=1}^{M}\left(f^2(x_k)-2f(x_k)\sum_{i=1}^{N}c_i\varphi_i(x_k)+\left(\sum_{i=1}^{N}c_i\varphi_i(x_k)\right)^2\right)$$ であるので，

$$\frac{\partial}{\partial c_l}E(c_1,\cdots,c_N)=\sum_{k=1}^{M}\left(-2f(x_k)\varphi_l(x_k)+2\varphi_l(x_k)\sum_{i=1}^{N}c_i\varphi_i(x_k)\right)=-2\sum_{k=1}^{M}f(x_k)\varphi_l(x_k)+2\sum_{i=1}^{N}\sum_{k=1}^{M}c_i\varphi_l(x_k)\varphi_i(x_k)=0$$

ここで，$(f,g)=\sum_{k=1}^{M}f(x_k)g(x_k)$と内積で表すと，$\sum_{i=1}^{N}c_i(\varphi_l,\varphi_i)=(f,\varphi_l)$が得られる．これが，$l=1,\cdots,N$について成り立つので，下記の連立方程式が得られ，これを解けば最適な係数が得られる．

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

$\sum_{k=1}^{M}\varphi_i(x_k)\varphi_j(x_k)=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$ という条件が満たされる場合には，上式の係数行列は単位行列となり，

$$\begin{bmatrix}c_1\\ \vdots\\ c_N\end{bmatrix}=\begin{bmatrix}(f,\varphi_1)\\ \vdots\\ (f,\varphi_N)\end{bmatrix}$$ が得られ,内積，すなわち直交展開係数が最適な係数となる．このことから，直交関数展開は最小二乗法でもあると言える．