統計解析特論 期末試験 予想問題

講義ノート，配布資料，その他の資料や本を持ち込み可とする (総重量は 1kg 以下とする)．電子機器は使用不可．
成績評価：レポート 50%，期末試験 50% として成績を評価する．

1. Ω を標本空間，事象 $A \subset \Omega$ の確率を $\Pr(A)$ とする．
   (a) 確率の公理を示せ．
   (b) 確率の公理から以下を導出せよ：
   $A, B \subset \Omega$ に対して $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
   (c) 確率の公理から以下を導出せよ：
   可算個の集合 $A_i \subset \Omega$, $i = 1, 2, 3, \ldots$ に対して $\Pr\big(\bigcup_i A_i\big) \le \sum_i \Pr(A_i)$.

2. 確率変数 X は，$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x \in \mathbb{R}$ を確率密度関数とする分布 (すなわち $N(0,1)$) にしたがうとする．
   (a) $Y = e^X$ の確率密度関数を求めよ．
   (b) Y の期待値を求めよ．

3. 事象 B の条件の下で事象 A が起こる条件付き確率を $\Pr(A|B)$ と表す．インフルエンザに感染している $(X = 1)$ か感染していない $(X = 0)$ か，また検査結果が陽性 $(Y = 1)$ か陰性 $(Y = 0)$ かの確率は，以下で与えられるとする．

$$\Pr(\, Y = 1 | X = 1\,) = p, \quad \Pr(\, Y = 1 | X = 0\,) = q, \quad \Pr(\, X = 1\,) = r.$$

   (a) $\Pr(X = 1 | Y = 1)$ の値を，p, q, r を用いて表せ．
   (b) $\Pr(X = 1 | Y = 0)$ の値を，p, q, r を用いて表せ．

4. 1 次元確率変数 Z の期待値と分散をそれぞれ $\mathbb{E}[Z], \mathbb{V}[Z]$ と表す．
   (a) $\mathbb{E}[Z^2] \ge \mathbb{V}[Z]$ を証明せよ．
   (b) $\mathbb{E}[Z^4] \ge (\mathbb{V}[Z])^2$ を証明せよ．
   (c) (b) の等号が成立する Z の例をひとつ示せ．ただし $\mathbb{V}[Z] \ne 0$ とする．

5. 確率変数 $Y_i, i = 1, \ldots, n$ は

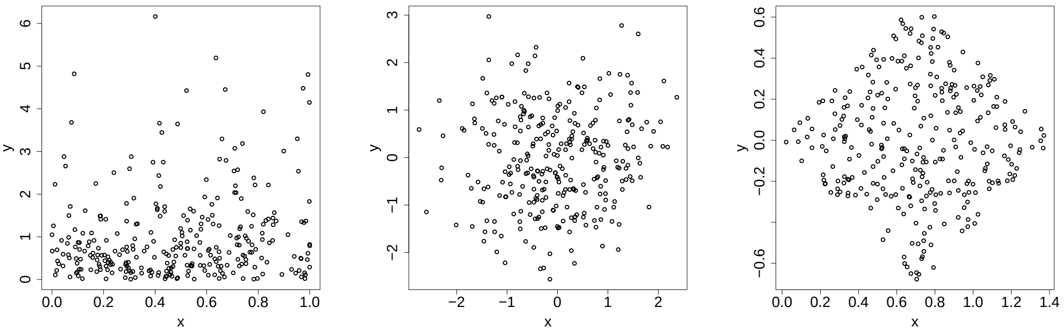
$$Y_i = \phi(x_i)^T \theta + \varepsilon_i,\ i = 1, \ldots, n, \quad \varepsilon_1 \ldots, \varepsilon_n \sim_{i.i.d.} N(0, \sigma^2)$$

   で定まる分布にしたがうとする．$x_1, \ldots, x_n$ は定数，また $\phi(x) \in \mathbb{R}^d, \theta \in \mathbb{R}^d$ とする．$n \times d$ 行列 Φ を

$$\Phi = (\phi(x_1), \ldots, \phi(x_n))^T$$

   とし，Φ のランク (階数) は d とする．
   (a) $Y = (Y_1, \ldots, Y_n)^T$ とする．パラメータ θ の最小 2 乗推定量を $\widehat{\theta}$ とする．
   (b) 点 x における関数値 $\phi(x)^T\theta$ を $\widehat{Y}(x) = \phi(x)^T\widehat{\theta}$ で予測する．$\widehat{Y}(x)$ の分散を求めよ．

6. 確率変数 (X,Y) の散布図 (a),(b),(c) を示す．



(a) (b) (c)

以下，X,Y の分布に関して正しいと考えられる選択肢を次の 1 〜 4 から選べ．また，その選択肢を正しいと考える理由も述べよ．

選択肢１．(a) は「独立でない」，(b) は「独立に異なる分布にしたがう」，(c) は「独立に同一の分布にしたがう」

選択肢２．(a) は「独立に異なる分布にしたがう」，(b) は「独立でない」，(c) は「独立に同一の分布にしたがう」

選択肢３．(a) は「独立に異なる分布にしたがう」，(b) は「独立に同一の分布にしたがう」，(c) は「独立でない」

選択肢４．(a) は「独立に同一の分布にしたがう」，(b) は「独立に異なる分布にしたがう」，(c) は「独立でない」

7. 線形独立なベクトル $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_d \in \mathbb{R}^n$ に対して $n \times d$ 行列を $\Phi = (\boldsymbol{a}_1,\ldots,\boldsymbol{a}_d)$ と定める．n 次正方行列 Π を $\Pi = \Phi(\Phi^T\Phi)^{-1}\Phi^T$ とする．

(a) $\mathrm{trace}(\Pi)$ を求めよ．

(b) Π の固有値と重複度 (固有空間の次元) をすべて求めよ．

8. $\mathcal{H}$ を RKHS, 対応するカーネル関数を $k(x,x')$, $f \in \mathcal{H}$ のノルムを $\|f\|$ と表す．データ $(x_1,y_1),\ldots,(x_n,y_n)$, $y_1,\ldots,y_n \in \mathbb{R}$ に対して，カーネル回帰関数を

$$\min_{f\in\mathcal{H}} \sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda\|f\|^2$$

によって推定する．ただし $\lambda > 0$ とし，上の最適化問題の最適解を $\widehat{f}_\lambda(x)$ とする．

(a) グラム行列 K を $K_{ij} = k(x_i,x_j), i,j = 1,\ldots n$ とし，$Y = (y_1,\ldots,y_n)^T \in \mathbb{R}^n$ とする．表現定理より $\widehat{f}_\lambda(x) = \sum_{i=1}^{n}\alpha_i k(x,x_i)$ と表せる．このとき $\alpha = (\alpha_1,\ldots,\alpha_n)^T \in \mathbb{R}^n$ を求めよ．

(b) $\widehat{f}_\lambda$ のノルム $\|\widehat{f}_\lambda\|$ は，λ の関数として単調減少かどうか調べよ．

9. $\mathcal{H}$ を RKHS，対応するカーネル関数を $k(x,x')$，また $f,g \in \mathcal{H}$ に対して内積とノルムをそれぞれ $\langle f,g\rangle, \|f\|$ と表す．

(a) 点 $x_1,\ldots,x_n$ に対して

$$\min_{f\in\mathcal{H}} \sum_{i=1}^{n}\|f - k(x_i,\cdot)\|^2$$

の最適解をひとつ求めよ．

(b) 点 $x_1,\ldots,x_n$ と定数 $\lambda > 0$ に対して

$$\min_{f\in\mathcal{H}} \lambda\|f\|^2 + \sum_{i=1}^{n}\|f - k(x_i,\cdot)\|^2$$

の最適解をひとつ求めよ．

10. $k_1(x,x'), k_2(x,x')$ がともに正定値カーネルなら，$k_1(x,x')k_2(x,x')$ も正定値カーネルであることを示せ．

11. $x, t \geq 0$ に対して定義される実数値関数 $f(x,t)$ に対して，次の積分値が定まるとする．

$$k(x,y) = \int_0^\infty f(x,t)f(y,t)dt.$$

(a) $k(x,y)$ は正定値カーネルであることを示せ．

(b) $0 \leq x, y \leq 1$ に対して $1 - \max\{x,y\}$ は正定値カーネルであることを示せ．

12. データ $(x_1,y_1),\dots(x_n,y_n) \in \mathbb{R}\times\mathbb{R}$ に対して線形モデル $y_i = \theta_0 + \theta_1 x_i + \varepsilon_i$, $\mathbb{E}[\varepsilon_i] = 0$ をあてはめるとき，最小２乗推定量を $\widehat{\theta}_0, \widehat{\theta}_1$ とする．一方，$a > 0, b, c \in \mathbb{R}$ を定数として，データ $(\widetilde{x}_i, \widetilde{y}_i) = (ax_i + b, y_i + c)$, $i = 1,\dots,n$ に対して線形モデル $\widetilde{y}_i = \theta_0 + \theta_1 \widetilde{x}_i + \widetilde{\varepsilon}_i$, $\mathbb{E}[\widetilde{\varepsilon}_i] = 0$ をあてはめるとき，最小２乗推定量を $\widetilde{\theta}_0, \widetilde{\theta}_1$ とする．$(\widehat{\theta}_0, \widehat{\theta}_1) = (-3, 1)$ のとき，$(\widetilde{\theta}_0, \widetilde{\theta}_1)$ を a, b, c を用いて表せ．

13. $\mathcal{H}$ を RKHS，対応するカーネル関数を $k(x,x')$，また $f, g \in \mathcal{H}$ に対して内積とノルムをそれぞれ $\langle f, g\rangle, \|f\|$ と表す．$\mathcal{H}$ 内の超平面を $P = \{f \in \mathcal{H} | \langle w, f\rangle + b = 0\}$ とする．ただし $w \in \mathcal{H}, b \in \mathbb{R}$, $w \neq 0$ とする．

(a) $f_0 \in \mathcal{H}$ に対して，$f_0 - \bar{f} = \alpha w$ を満たす $\bar{f} \in P$ と $\alpha \in \mathbb{R}$ を選ぶ．このとき，$\bar{f}$ は

$$\min_{f \in P} \|f_0 - f\|$$

の最適解であることを示せ．

(b) $\min_{f\in P} \|f_0 - f\|$ の値と α をそれぞれ求めよ．

14. 独立に同一の分布にしたがう２値データを $D = \{(x_1,y_1),\dots,(x_n,y_n)\}$, $y_1,\dots,y_n \in \{+1,-1\}$ とする．仮説集合 $\mathcal{H}$ のなかで，学習誤差 $\widehat{e}(h) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}[h(x_i) \neq y_i]$ を最小にする仮説を $\widehat{h}$，また 予測誤差 $e(h) = \mathbb{E}[\mathbf{1}[h(x) \neq y]]$ を最小にする仮説を $h_{\mathcal{H}}$ とする．以下，$\mathbb{E}_D[\,\cdot\,]$ を観測データ D の分布に関する期待値とするとき，以下の不等式が成り立つことを示せ．

$$\mathbb{E}_D[\widehat{e}(\widehat{h})] \;\leq\; e(h_{\mathcal{H}}) \;\leq\; \mathbb{E}_D[e(\widehat{h})].$$

15. ２値データ (x,y), $y \in \{+1,-1\}$ に対して，判別関数 $f(x)$ の２乗損失は

$$\ell(yf(x)) = (y - f(x))^2$$

で与えられる．期待損失

$$\mathbb{E}[\ell(yf(x))] = \int \sum_{y=\pm 1} (y - f(x))^2 \mathrm{Pr}(Y = y|x)p(x)dx$$

を最小にする判別関数 $f(x)$ を求めよ．また $\mathrm{sign}(f(x))$ がベイズルールに一致するかどうか調べよ．

16. $Z_1,\dots,Z_n$ は区間 $[0,1]$ に値をとる確率変数とし，$Z_1,\dots,Z_n \sim_{i.i.d.} P$, $\mathbb{E}[Z_i] = \mu$ とする．区間 $(0,1)$ に値をとる狭義単調減少関数 $\gamma(\cdot)$ が存在して

$$\text{(確率不等式)} \qquad \mathrm{Pr}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i - \mu\right| \geq \varepsilon\right) \leq \gamma(n\varepsilon^2)$$

が成り立つと仮定する．$\gamma(\cdot)$ の逆関数を $\gamma^{-1}(\cdot)$ とする．この不等式を用いて，以下の問に答えよ．

(a) $\delta \in (0,1)$ に対して $\mathrm{Pr}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i - \mu\right| \geq \varepsilon\right) \leq \delta$ が成り立つように ε を定めよ．

(b) ２値判別において，n 個のデータがある分布に独立にしたがうとする．損失は 0-1 損失として，仮説 h の学習誤差を $\widehat{e}(h)$，予測誤差を $e(h)$ とする．有限個の仮説からなる仮説集合 $\mathcal{H}$ に対して，学習誤差の最小化 (ERM) により仮説 $\widehat{h}$ を学習するとき，$1 - \delta$ 以上の確率で

$$e(\widehat{h}) < \widehat{e}(\widehat{h}) + \varepsilon$$

が成り立つとする．上の確率不等式を用いて，ε を $\delta, n, |\mathcal{H}|$ で表せ．ただし，$|\mathcal{H}|$ は $\mathcal{H}$ の要素数とする．

17. k-重交差検証法 (k-fold cross validation) の目的と手法を簡潔に説明せよ．

18. 判別問題におけるベイズ誤差について，以下の問に答えよ．ただし，損失は 0-1 損失とする．

    (a) 2 値判別問題では，ベイズ誤差は 1/2 以下であることを示せ．

    (b) ラベル集合 $\mathcal{Y} = \{1, 2, \ldots, G\}$, $G \geq 3$ に対する多値判別問題では，ベイズ誤差は $(G-1)/G$ 以下であることを示せ．

19. 誤り訂正出力符号法 (ECOC) で多値判別を行う．ラベル $\mathcal{Y} = \{1, 2, 3, 4\}$ に対して，以下のように $c_1, c_2, c_3, c_4 \in \{+1, 0, -1\}^T$ を定める．

	t $(T=5)$				
label	1	2	3	4	5
1 (c_1)	-1	-1	-1	-1	0
2 (c_2)	$+1$	$+1$	-1	0	-1
3 (c_3)	0	-1	$+1$	$+1$	-1
4 (c_4)	$+1$	0	-1	$+1$	$+1$

2 つの符号 $c = (c_1, \ldots, c_T)$, $c' = (c'_1, \ldots, c'_T) \in \{+1, 0, -1\}^T$ の違いを $d(c, c') = \displaystyle\sum_{t=1}^{T} \frac{1 - c_t c'_t}{2}$ で測る．

    (a) 上の表において $d(c_1, c_2)$ を求めよ．

    (b) ECOC で仮説 $h_t(x)$, $t = 1, \ldots, 5$ を学習したとき, 入力 x に対して符号 $(h_1(x), \ldots, h_5(x)) = (+1, +1, +1, -1, -1)$ が得られたとする．予測ラベルを求めよ．

    (c) 上の表において $\rho = \min_{y \neq y'} d(c_y, c_{y'})$ を求めよ．

    (d) 上の表に $t = 6$ の列を追加し，$T = 6$ の符号化法に拡張する．拡張した符号化法の $\rho = \min_{y \neq y'} d(c_y, c_{y'})$ が最も大きくなるように，$t = 6$ の列を定めよ．(解は複数あるが，そのうち 1 つを挙げればよい)

20. ラベルを $\mathcal{Y} = \{1, \ldots, G\}$ として，誤り訂正出力符号法 (ECOC) で多値判別を行う．ラベル y に符号 $c_y = (c_{y1}, \ldots, c_{yT}) \in \{+1, 0, -1\}^T$ をランダムに割り当てる．具体的には，各 c_{yt}, $y \in \mathcal{Y}, t = 1, \ldots T$ は独立に同一の分布

$$\Pr(c_{yt} = +1) = p, \quad \Pr(c_{yt} = -1) = q, \quad \Pr(c_{yt} = 0) = r$$

によって定まるとする．ここで $p + q + r = 1$ である．2 つの符号 $c = (c_1, \ldots, c_T)$, $c' = (c'_1, \ldots, c'_T) \in \{+1, 0, -1\}^T$ の違いを $d(c, c') = \displaystyle\sum_{t=1}^{T} \frac{1 - c_t c'_t}{2}$ で測る．以下，異なる 2 つのラベルを $y, y' \in \mathcal{Y}$ とする．

    (a) $d(c_y, c_{y'})$ の期待値 $\mathbb{E}[d(c_y, c_{y'})]$ を求めよ．

    (b) $\mathbb{E}[d(c_y, c_{y'})]$ を p, q の関数とする．$\mathbb{E}[d(c_y, c_{y'})]$ を最大にする p, q の条件を求めよ．

    (c) $0 \leq r \leq 1/2$ を仮定する．(b) の条件のもとで分散 $\mathbb{V}[d(c_y, c_{y'})]$ を最小にする p, q, r を求めよ．

21. $\mathcal{X}$ を集合，関数 $k(x, x')$, $x, x' \in \mathcal{X}$ を $k(x, x') = \mathbf{1}[x = x']$ と定める．

    (a) $k(x, x')$ は正定値カーネルであることを証明せよ．

    (b) 2 値データ $(x_1, y_1), \ldots, (x_n, y_n)$, $x_i \in \mathcal{X}$, $y_i \in \{+1, -1\}$ が与えられたとする．ここで x_i はすべて異なる点とする (すなわち $i \neq j$ なら $x_i \neq x_j$)．関数 $k(x, x') = \mathbf{1}[x = x']$ を用いたカーネル SVM で判別関数を学習する：

$$\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \max\{1 - y_i f(x_i), 0\} + \frac{\lambda}{2} \|f\|^2.$$

    ここで $\mathcal{H}$ は $k(x, x')$ に対応する RKHS, $\|f\|$ は $k(x, x')$ から定まる $\mathcal{H}$ のノルムとする．上の最適化問題の最適解 $\widehat{f}(x)$ を求めよ．また，$\widehat{f}(x)$ の学習誤差 $\widehat{e}(\widehat{f}) = \displaystyle\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}[\widehat{f}(x_i) \neq y_i]$ を求めよ．