דף נוסחאות למבחן

הגדרות בסיסיות בהסתברות

תוחלת (mean/expectation) של משתנה אקראי:

$$E[X]=\sum_i x_i p(x_i)\equiv \mu$$

שונות (variance) של משתנה אקראי:

$$Var(X)=E[(X-\mu)^2]=E[X^2]-\mu^2\equiv\sigma^2$$

שונות משותפת (covariance) של שני משתנים אקראיים X ו Y:

$$Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]=E[XY]-\mu_X\mu_Y$$

ההסתברות המותנית של Y בהינתן X מוגדרת ע"י:

$$p(y\,|\,x)=\frac{p(x,y)}{p(x)}$$

ההסתברות המותנית של Y בהינתן X עבור התניה במשתנה שלישי Z:

$$p(x,y\,|\,z)=p(y\,|\,x,z)\,p\left(x\,|\,z\right)$$

כלל Bayes:

$$p(y\,|\,x)=\frac{p(x\,|\,y)p(y)}{p(x)}=\frac{p(x\,|\,y)p(y)}{\sum_i p(x\,|\,y_i)p(y_i)}$$

צפיפות ההתפלגות של וקטור אקראי גאוסי:

$$f(x_1,x_2...,x_k)=\frac{1}{\sqrt{(2\pi)^k|\mathbf{\Sigma}|}}\exp\left(-\frac{1}{2}\left(\mathbf{X}-\boldsymbol{\mu}\right)^T\mathbf{\Sigma}^{-1}\left(\mathbf{X}-\boldsymbol{\mu}\right)\right)$$

חוקי גזירה וקטורית

$\frac{\partial}{\partial\underline{x}}\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)^T\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)=2\underline{\underline{A}}^T\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)$
$\frac{\partial}{\partial\underline{x}}\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)^T\underline{\underline{C}}\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)=\underline{\underline{A}}^T\left(\underline{\underline{C}}+\underline{\underline{C}}^T\right)\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)$
$\frac{\partial}{\partial\underline{x}}\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)^T\left(\underline{\underline{D}}\underline{x}+\underline{e}\right)=\underline{\underline{A}}^T\left(\underline{\underline{D}}\underline{x}+\underline{e}\right)+\underline{\underline{D}}^T\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)$
$\frac{\partial}{\partial\underline{x}}\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)^T\underline{\underline{C}}\left(\underline{\underline{D}}\underline{x}+\underline{e}\right)=\underline{\underline{A}}^T\underline{\underline{C}}\left(\underline{\underline{D}}\underline{x}+\underline{e}\right)+\underline{\underline{D}}^T\underline{\underline{C}}^T\left(\underline{\underline{A}}\underline{x}+\underline{b}\right)$

$\frac{\partial}{\partial\underline{x}}\left(\underline{a}^T\underline{x}\right)=\frac{\partial}{\partial\underline{x}}\left(\underline{x}^T\underline{a}\right)=\underline{a}$
$\frac{\partial}{\partial\underline{x}}\left(\underline{x}^T\underline{\underline{A}}\underline{x}\right)=\left(\underline{\underline{A}}^T+\underline{\underline{A}}\right)\underline{x}$
$\frac{\partial z}{\partial\underline{x}}=\frac{\partial\underline{\underline{y}}}{\partial\underline{x}}\frac{\partial z}{\partial\underline{y}}$
$\frac{\partial z}{\partial\underline{x}}\neq\frac{\partial z}{\partial\underline{y}}\frac{\partial\underline{\underline{y}}}{\partial\underline{x}}$ בדרך כלל
$\frac{\partial}{\partial\underline{x}}\underline{\underline{A}}^T\underline{x}=\frac{\partial}{\partial\underline{x}}\underline{x}^T\underline{\underline{A}}=\underline{\underline{A}}$

<u>שערוך פרמטרים</u>

משערך הסבירות המירבית (Maximum Likelihood Estimator) של θ:

$$\hat{\theta}_{MLE} \triangleq \underset{\theta\in\mathbb{R}^p}{\arg\max}\, p(D\mid\theta) = \underset{\theta\in\mathbb{R}^p}{\arg\max}\left\{\log p(D\mid\theta)\right\}$$

משערך MAP:

$$\hat{\theta}_{MAP} \triangleq \underset{\theta\in\mathbb{R}^p}{\arg\max}\, p(\theta\mid D) = \underset{\theta\in\mathbb{R}^p}{\arg\max}\left\{\log p(\theta\mid D)\right\}$$

<u>סיווג בייסיאני</u>

$$f(x) = \underset{i=1,\ldots,N}{\arg\max}\, p(y=C_i\mid x)$$

<u>פונקציית הרגרסיה הלינארית</u>

מנסים למצוא קירוב טוב ככל האפשר ל- $f_0(x) = Y-\varepsilon$ ע"י

$$\hat{f}(x,w) = w^T\phi(x) = \sum_{m=1}^{M} w_m\phi_m(x)$$

כאשר $\left\{\phi_m(x)\right\}_{m=1}^{M}$ הוא אוסף פונקציות בסיס שנבחרות מראש, ו- $w=\left(w_1,\ldots,w_M\right)^T$ הוא ווקטור הפרמטרים הנלמדים, $w\in\mathbb{R}^M$.

בחירת הפרמטרים תתבצע לפי קריטריון **השגיאה האמפירית הריבועית** ביחס לסדרת הלימוד:

$$w^* = \underset{w\in\mathbb{R}^M}{\arg\min}\left\{\sum_{k=1}^{n}\left(y_k-\hat{f}\left(x_k,w\right)\right)^2\right\} = \underset{w\in\mathbb{R}^M}{\arg\min}\left\{\sum_{k=1}^{n}\left(y_k-w^T\phi(x_k)\right)^2\right\} \quad (*)$$

<u>קריטריונים לגידול עץ החלטה:</u>

1. שגיאת הסיווג: $Q(S) = 1-\max_{j\in\{1,\ldots,C\}}\hat{p}_j$
2. אינדקס Gini: $Q(S) = \sum_j \hat{p}_j(1-\hat{p}_j)$
3. אנטרופיה: $Q(S) = H(S) = \sum_j \hat{p}_j\log_2\frac{1}{\hat{p}_j} = -\sum_j \hat{p}_j\log_2(\hat{p}_j)$

<u>אופטימיזציה ללא אילוצים</u>

נתונה פונקציית מחיר, $f(\boldsymbol{x}):\mathbb{R}^n \to \mathbb{R}$, התלויה בווקטור פרמטרים $\boldsymbol{x}=(x_1,x_2,\dots,x_n)^T$.

נרצה למצוא $\boldsymbol{x}^*$ המביא למינימום את פונקציית המחיר, $f(\boldsymbol{x}^*)\le f(\boldsymbol{x}) \quad \forall \boldsymbol{x}\in X$, כאשר אין אילוצים על $\boldsymbol{x}$.

<u>תנאים לאופטימליות</u>

היעד: $\mathbf{w}^* = \arg\min_{w\in\mathbb{R}^n} f(\mathbf{w})$.

בהנחה ש- f גזירה, תנאי הכרחי לנקודת מינימום מקומית הוא: $\nabla f(\mathbf{w}^*)=0$

אם בנוסף f גזירה פעמיים, תנאי מספיק לנקודת מינימום מקומית מבודדת הוא:

$$\mathbf{H}(\mathbf{w}^*)>0 \text{ and } \nabla f(\mathbf{w}^*)=0$$

משמעות אי-השוויון היא שההסיאן הוא מטריצה חיובית מוגדרת (שכל הערכים העצמיים שלה חיוביים).

<u>אלגוריתם הגרדיאנט לאופטימיזציה נטולת אילוצים:</u>

מטרה: במקרים רבים לא ניתן לחשב את $\mathbf{w}^*$ באופן ישיר ולכן משתמשים <u>בטכניקות פתרון איטרטיביות</u>. כלומר, מתחילים מניחוש $\mathbf{w}(0)=\mathbf{w}_0$ ומייצרים סדרה $\mathbf{w}(t+1)=\mathbf{w}(t)+\Delta\mathbf{w}(t)$, כך שיתקיים $f(\mathbf{w}(t+1))\le f(\mathbf{w}(t))$, "בתקווה" שהסדרה תתכנס למינימום.

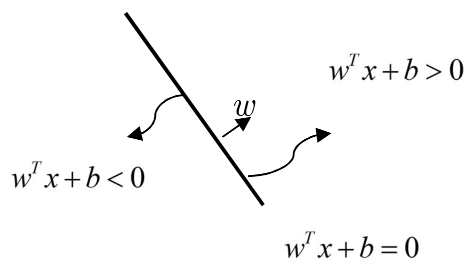
קלט: רמת דיוק $thresh$ ונקודת מוצא $\mathbf{w}_0$.

האלגוריתם:

1. אתחל את נקודת המוצא של האלגוריתם $\mathbf{w}_0$.
2. עדכן את הפרמטרים בכיוון הנגדי לגרדיאנט בנקודה:
$$\mathbf{w}_{t+1}=\mathbf{w}_t+\Delta\mathbf{w}_t$$
$$\Delta\mathbf{w}_t=-\eta\mathbf{g}(\mathbf{w}_t)$$
3. בדיקת התכנסות:
   - אם $|\mathbf{w}_{t+1}-\mathbf{w}_t|\le thresh$
   סיים אלג' ותן נקודה $\mathbf{w}_{t+1}$ כפתרון.
   - אחרת – חזור לשלב 2.

גאומטריה של מישור

- היטל של וקטור x בכיוון וקטור w: $\frac{w^T x}{\|w\|}$
- משוואה של מישור ב $\mathbb{R}^d$: $w^T x + b = 0$, עבור $w \in \mathbb{R}^d, b \in \mathbb{R}$ קבועים המגדירים את המישור. מהמשוואה ברור ש cw, cb עבור קבוע כלשהוא c מגדירים את אותו מישור כמו w, b.
- מרחק אוקלידי של נקודה x_0 מהמישור המוגדר ע"י w, b: $d = \frac{w^T x_0 + b}{\|w\|}$. בהצבת $x = 0$ מתקבל שהמרחק של הראשית מהמישור הוא $d_0 = \frac{b}{\|w\|}$. הסימן של d קובע האם הנקודה בצד של המישור לכיוון המקביל ל w, או לכיוון האנטי-מקביל ל w:



SVM

ניסוח הבעיה הפרימאלית: (P)

$$\boxed{\begin{aligned} &\min_{w,b} \tfrac{1}{2}\|w\|^2 \\ &\text{s.t.}: \quad y_k(w^T x_k + b) \geq 1, \quad k = 1,2,\ldots,n \end{aligned}}$$

ניסוח הבעיה הדואלית: (D)

$$\boxed{\begin{aligned} &\max_{\alpha} \sum_{k=1}^{n} \alpha_k - \frac{1}{2}\sum_{k,l=1}^{n} \alpha_k \alpha_l y_k y_l \langle x_k, x_l \rangle \\ &\text{s.t.}: \quad \alpha_k \geq 0, \quad k = 1,2,\ldots,n \\ &\textstyle\sum_{k=1}^{n} \alpha_k y_k = 0 \end{aligned}}$$

<u>חסמי PAC:</u>

<u>עבור מחלקת היפותזות סופית :</u>

- חסם למידה אגנוסטית ($f^* \notin F$):

$$(1)\quad \forall \varepsilon > 0: \ \mathbb{P}\left\{L(\hat{f}_n) - L^* > \varepsilon\right\} < \ 2\left|F\right| e^{-\varepsilon^2 n/2}$$

- חסם על שגיאת השערוך עבור חזאי בר-מימוש ($f^* \in F$):

$$(2)\quad \forall \varepsilon > 0: \ \mathbb{P}\left\{L(\hat{f}_n) > \varepsilon\right\} < \left|F\right| e^{-\varepsilon n}$$

<u>עבור מחלקת היפותזות אינסופית :</u>

- עבור מקדם ניתוץ ידוע:

$$\forall \varepsilon > 0: \ \mathbb{P}\left\{L(\hat{f}_n) - L^* > \varepsilon\right\} < \ 4 S_F\left(2n\right) e^{-\varepsilon^2 n/32}$$

- עבור $VC\dim = V$:

$$\forall \varepsilon > 0: \ \mathbb{P}\left\{L(\hat{f}_n) - L^* > \varepsilon\right\} < \ 4(2n+1)^V e^{-\varepsilon^2 n/32}$$

<u>מימד VC</u>

- ממד VC של מחלקת היפותזות F נתון ע"י:

$$\begin{aligned} VC(F) &\triangleq \max\{n \ge 1 \ : \ S_F(n) = 2^n\} \\ &\equiv \max\{n \ge 1 \ : \ \max_{\{x_1 \dots x_n\} \subset X} S_F\{x_1 \dots x_n\} = 2^n\} \end{aligned}$$

- מקדם הניתוץ מוגדר ע"י המקסימום של $S_F\{x_1,\dots,x_n\}$ על כל קבוצות הנקודות האפשריות, כלומר:

$$S_F(n) \triangleq \max_{\{x_1 \dots x_n\} \subset X} S_F\{x_1,\dots,x_n\}$$

- הלמה של Sauer:

$$S_F(n) \le \sum_{i=0}^{V} \binom{n}{i} \le \left(n+1\right)^V$$

- $S_F\{x_1,\dots,x_n\} \triangleq |\{(f(x_1),\dots,f(x_n)) : f \in F\}|$ מייצג את כמות הדיכוטומיות השונות שמחלקה F משרה על הקבוצה $\{x_1,\dots,x_n\}$.

<u>נורמות:</u>

עבור וקטור $x \in \mathbb{R}^m$

נורמת L_1:	$\|x\|_1 = \sum_{i=1}^{n} \|x_i\|$
נורמת L_2:	$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$
נורמת L_∞:	$\|x\|_\infty = \max_i x_i$
נורמת L_p:	$\|x\|_p = \left(\sum_{i=1}^{n} \|x_i\|^p \right)^{1/p}$