תורת ההחלטות הביאסיאנית

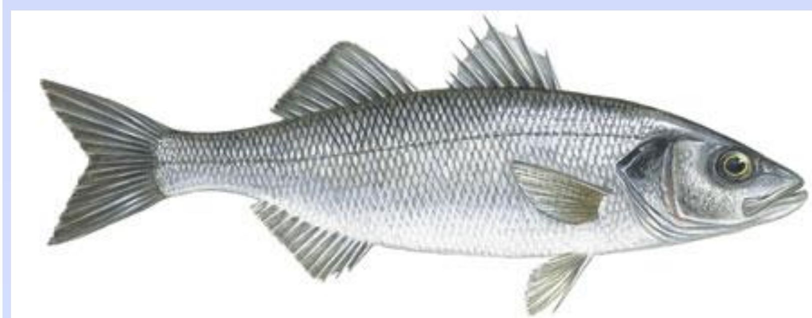
- הגישה מבוססת על כימות הפשרות בין החלטות סווג שונות תוך שימוש בהסתברות ובעלות של ההחלטות
- הנחות:

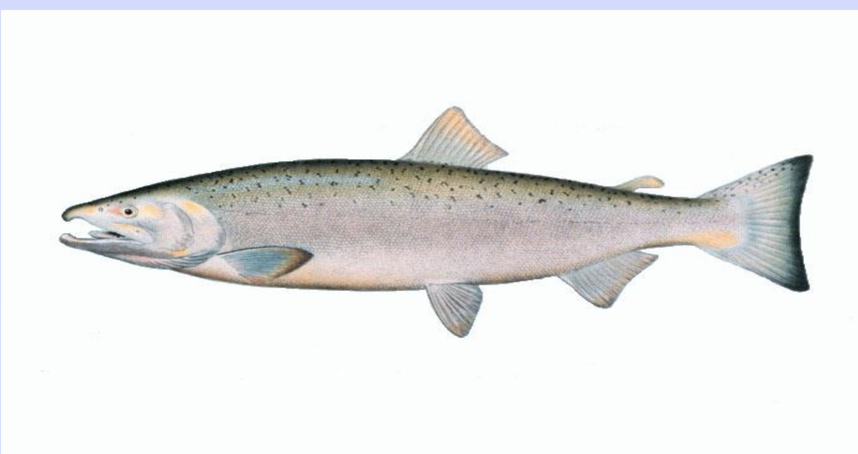
1. בעיית ההחלטה ניתנת לתיאור במונחים הסתברותיים
2. כל הערכים ההסתברותיים (צפיפויות הסתברות) ידועים

תורת ההחלטות הביאסיאנית

דוגמא

- נדון בבעיית הסווג של שני סוגי דגים: לברק (Sea bass) וסלמון.
- הדגים מופיעים באופן אקראי לפי שכיחותם





תורת ההחלטות הביאסינית

- נסמן: ω - מצב הטבע

$$\omega = \omega_1 - Sea\ \ bass$$

$$\omega = \omega_2 - Salmon$$

ω - משתנה אקראי

תורת ההחלטות הביאסנית

$$p(\omega_1) - the\ next\ \ fish\ is\ a\ Sea\ bass$$

$$p(\omega_2) - the\ next\ \ fish\ is\ a\ Salmon$$

- מצב הטבע הוא משתנה אקראי
- הסתברות (אפריורית) שווה ללכוד לברק או סלמון

$P(\omega_1) = P(\omega_2)$ (uniform priors)

אם אין דגים אחרים :

$P(\omega_1) + P(\omega_2) = 1$ (מאורעות זרים ומשלימים)

תורת ההחלטות הביאסינית

- מהו כלל ההחלטה ללא מידע קודם ?
  - Decide - ω_1 if $P(\omega_1) > P(\omega_2)$

    - otherwise decide ω_2

    (ההנחה היא שלכל החלטה יש אותה עלות (cost))

    שאלה: האם מוצדק להשתמש בכלל הזה?

תורת ההחלטות הביאסיאנית

- אם $P(\omega_1) >> P(\omega_2)$, השגיאה תהיה קטנה ורוב הזמן נחליט נכון.
- ברור שאם $P(\omega_1) \approx P(\omega_2)$ תהיה אפשרות של שגיאה של 50%.

בהמשך נראה כי ההסתברות לשגיאה היא $\min(P(\omega_1) , P(\omega_2))$ (בנסיבות אלה אין כלל החלטה שישיג הסתברות גבוהה יותר להחלטה נכונה).

תורת ההחלטות הביאסיאנית

- *כשיש מידע נוסף: כדי לבצע החלטה יותר נכונה (המתבססת על המידע) מודדים (למשל) את בהירות הדג:*
- משתמשים במידע עבור כל מחלקה
- $P(x \mid \omega_1)$ and $P(x \mid \omega_2)$ מתארות את ההבדלים בבהירות בין שתי האוכלוסיות: אוכלוסיית הסלמון ואוכלוסיית הלברק. (difference in lightness)
- $P(x|\omega_i)$ – *class or state conditional pdf*
- x – משתנה אקראי רציף, שהתפלגותו תלוייה במצב הטבע

תורת ההחלטות הביאסינית

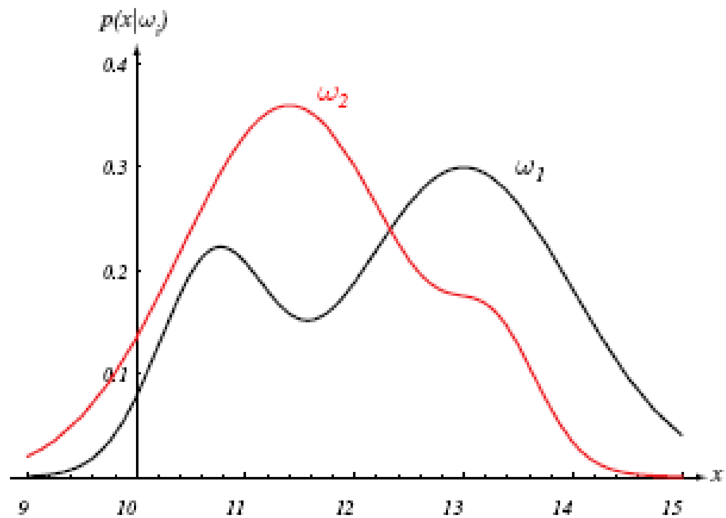


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

צפיפות ההסתברות של x בהינתן המחלקה ω_i . *אם x מייצג את הבהירות של הדג, העקומות מתארות את ההבדל בבהירות של שני סוגי הדגים. פונקציות הצפיפות מנורמלות, כך שהשטח מתחת לכל עקומה הוא 1.0.*

תורת ההחלטות הביאסינית

- הנחה: ההסתברויות $P(x \mid \omega_1)$ ו- $P(x \mid \omega_2)$

וכן $P(\omega_1)$, $P(\omega_2)$ ידועות.

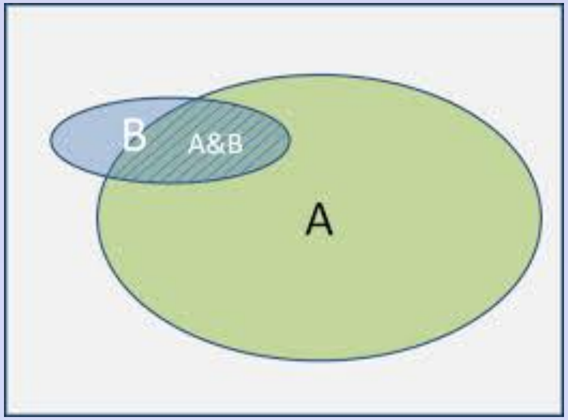
- נניח שמבצעים מדידה של x.

כיצד משפיעה המדידה על ההחלטה?

תזכורת

$$P(A|B)=\frac{P(A\bigcap B)}{P(B)}=\frac{P(A,B)}{P(B)}$$

$$P(B|A)=\frac{P(A\bigcap B)}{P(A)}=\frac{P(A,B)}{P(A)}$$



$$P(A,B)=P(A\bigcap B)=P(A|B)P(B)=P(B|A)\cdot P(A)$$

נוסחת Bayes:

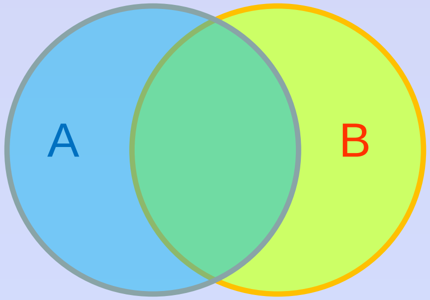
$$P(A|B)=\frac{P(B|A)\cdot P(A)}{P(B)}$$

תורת ההחלטות הביאסינית

- Posterior, likelihood, evidence

$$P(\omega_j, x) = P(\omega_j | x) \cdot P(x) = P(x | \omega_j) \cdot P(\omega_j)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A,B)}{P(B)}$$



$$P(\omega_j | x) = \frac{P(x|\omega_j) \cdot P(\omega_j)}{P(x)}$$

– במקרה של שתי מחלקות:

$$P(x) = \sum_{j=1}^{2} P(x|\omega_j) \cdot P(\omega_j)$$

(נוסחת ההסתברות השלמה)

תורת ההחלטות הביאסינית

Posterior = (Likelihood · Prior) / Evidence

– נוסחת Bayes

$$P(\omega_j \mid x) = \frac{\overbrace{P(x|\omega_j)}^{likelihood} \cdot \overbrace{P(\omega_j)}^{prior}}{\underbrace{\sum_j P(x|\omega_j) \cdot P(\omega_j)}_{evidence}}$$

ההסתברות של מצב
הטבע בהינתן המדידה x

נגדיר: הסתברות אפוסטריורית – $P(\omega_j|x)$

הסתברות אפריורית – $P(\omega_j)$

הסבירות של ω_j ביחס ל- x $P(x|\omega_j)$

בהינתן שכל שאר הגורמים שווים, הקטיגוריה
עבורה הסבירות היא הגבוהה ביותר, סביר
להניח שהיא הקטיגוריה האמיתית.

תורת ההחלטות הביאסינית

$$P(\omega_j \mid x) = \frac{P(x|\omega_j) \cdot P(\omega_j)}{\sum_j P(x|\omega_j) \cdot P(\omega_j)}$$

תורת ההחלטות הביאסיאנית

• נשים לב כי המכפלה $P(\omega_j) \cdot P(x|\omega_j)$ היא החשובה, כי המכנה p(x) זהה לכל החישובים – גורם נירמול המבטיח שסכום ההסתברויות יהיה שווה ל- 1.

$$\sum_j P(\omega_j | x) = 1$$

תורת ההחלטות הביאסינית

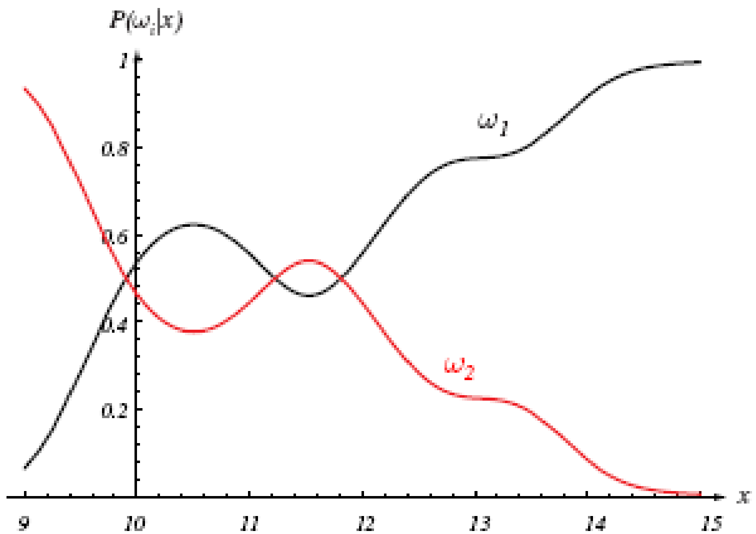


FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*.

$$P(\omega_1) = \frac{2}{3}$$

$$P(\omega_2) = \frac{1}{3}$$

לדוגמא בהינתן שהתבנית נמדדה ו – x=14 ההסתברות שהתבנית היא בקטיגוריה 1 היא ~0.92 ובקטיגוריה 2 היא ~0.08 . עבור כל X סכום ההסתברויות האפוסטריוריות הוא 1.

תורת ההחלטות הביאסינית

- ההחלטה בהינתן ההסתברויות האפוסטריוריות:

- עבור התצפית x:
- if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ ⇨ True state of nature = ω_1
- if $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ ⇨ True state of nature = ω_2

- לכן:
- עבור כל תצפית של x מסויים, ההסתברות לשגיאה היא:
- $P(error \mid x) = P(\omega_1 \mid x)$ *if we decide* ω_2
- $P(error \mid x) = P(\omega_2 \mid x)$ *if we decide* ω_1

הסבר: אם מחליטים ω2 מה הסיכוי לשגיאה?

זוהי ההסתברות שמצב הטבע הוא ω1 בהינתן התצפית x , כלומר $P(\omega_1 \mid x)$

תורת ההחלטות הביאסיאנית

איך נמזער את השגיאה עבור X נתון?

- Minimizing the probability of error

- Decide ω_1 *if* $P(\omega_1 \mid x) > P(\omega_2 \mid x)$; otherwise decide ω_2

תורת ההחלטות הביאסיאנית

ברור שמדידה מסויימת X עשוייה להיות חד-פעמית, ולכן נשאל, בהינתן כלל ההחלטה הנ"ל:

- (Decide ω_1 *if* $P(\omega_1 \mid x) > P(\omega_2 \mid x)$; otherwise decide ω_2)

האם ממזערים גם את השגיאה הממוצעת?

ההסתברות הממוצעת לשגיאה:

$$\int_{-\infty}^{\infty} p(error, x)dx = \int_{-\infty}^{\infty} p(error \mid x)\, p(x)dx$$

התשובה היא כן, כי אם מבטיחים שההסתברות לשגיאה תהיה מינימלית עבור כל x, אזי גם האינטגרל חייב להיות קטן ככל האפשר.

תורת ההחלטות הביאסינית

- Decide ω_1 *if* $P(\omega_1 \mid x) > P(\omega_2 \mid x)$; otherwise decide ω_2

- השגיאה הופכת להיות:

$$P(error \mid x) = \min\,[P(\omega_1 \mid x), P(\omega_2 \mid x)]$$

(Bayes decision)

כלל ההחלטה מדגיש את חשיבות ההסתברות האפוסטריורית.

תורת ההחלטות הביאסיאנית

על-ידי שימוש בכלל Bayes

$$P(\omega_j \mid x) = \frac{P(x|\omega_j) \cdot P(\omega_j)}{\sum_j P(x|\omega_j) \cdot P(\omega_j)}$$

והשמטת p(x) (שכאמור הוא גורם scaling בלבד),

נקבל כלל החלטה שקול:

- Decide ω_1 *if* $P(x|\omega_1)\,P(\omega_1) > P(x|\omega_2)\,P(\omega_2)$ otherwise decide ω_2

תורת ההחלטות הביאסיאנית

מקרים פרטיים:

1) $P(x|\omega_1) = P(x|\omega_2)$ במקרה זה הסבירות אינה נותנת מידע וההחלטה היא על סמך ההסתברויות האפריוריות בלבד.

2) אם $P(\omega_1) = P(\omega_2)$ במקרה זה כלל ההחלטה הוא על סמך הסבירויות (likelihoods) בלבד.

תורת ההחלטות הביאסינית

תכונות רציפות

- הכללת הרעיונות הקודמים
  - שימוש ביותר מתכונה אחת
  - נשתמש ביותר משתי מחלקות
  - נאפשר פעולות ולא רק החלטות על מצב הטבע
  - נציג פונקציית עלות כללית יותר מפונקציית השגיאה

תורת ההחלטות הביאסיאנית

- האפשרות של פעולות נוספות (פרט להחלטות סווג) מאפשר (למשל) לדחות החלטה
- סירוב לבצע החלטה עבור דוגמאות קרובות או גרועות
- פונקציית העלות או ההפסד אומרת מה העלות או המחיר של כל פעולה

תורת ההחלטות הביאסיאנית

נפרט: במקום הסקלר x נשתמש בוקטור תכונות $\underline{x}$, כאשר וקטור התכונות הוא במרחב תכונות d –מימדי

$$\underline{x} \in R^d$$

יותר משתי קטיגוריות – שימושי יותר וכללי יותר

יותר פעולות – בעיקר הכוונה לאפשרות לא להחליט

פונקציות עלות או מחיר כלליות – קובעות בדיוק מה העלות של כל פעולה

תורת ההחלטות הביאסיאנית

- נניח כי $\{\omega_1, \omega_2, \ldots, \omega_c\}$ קבוצת C המחלקות (קבוצה סופית)
- נניח כי $\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$ היא קבוצת a הפעולות (1,2,…,a)

- נסמן ב- $\lambda(\alpha_i \mid \omega_j)$ את העלות של נקיטת הפעולה α_i כאשר המצב האמיתי הוא ω_j

תורת ההחלטות הביאסינית

חישוב ההסתברות האפוסטריורית על-ידי כלל Bayes:

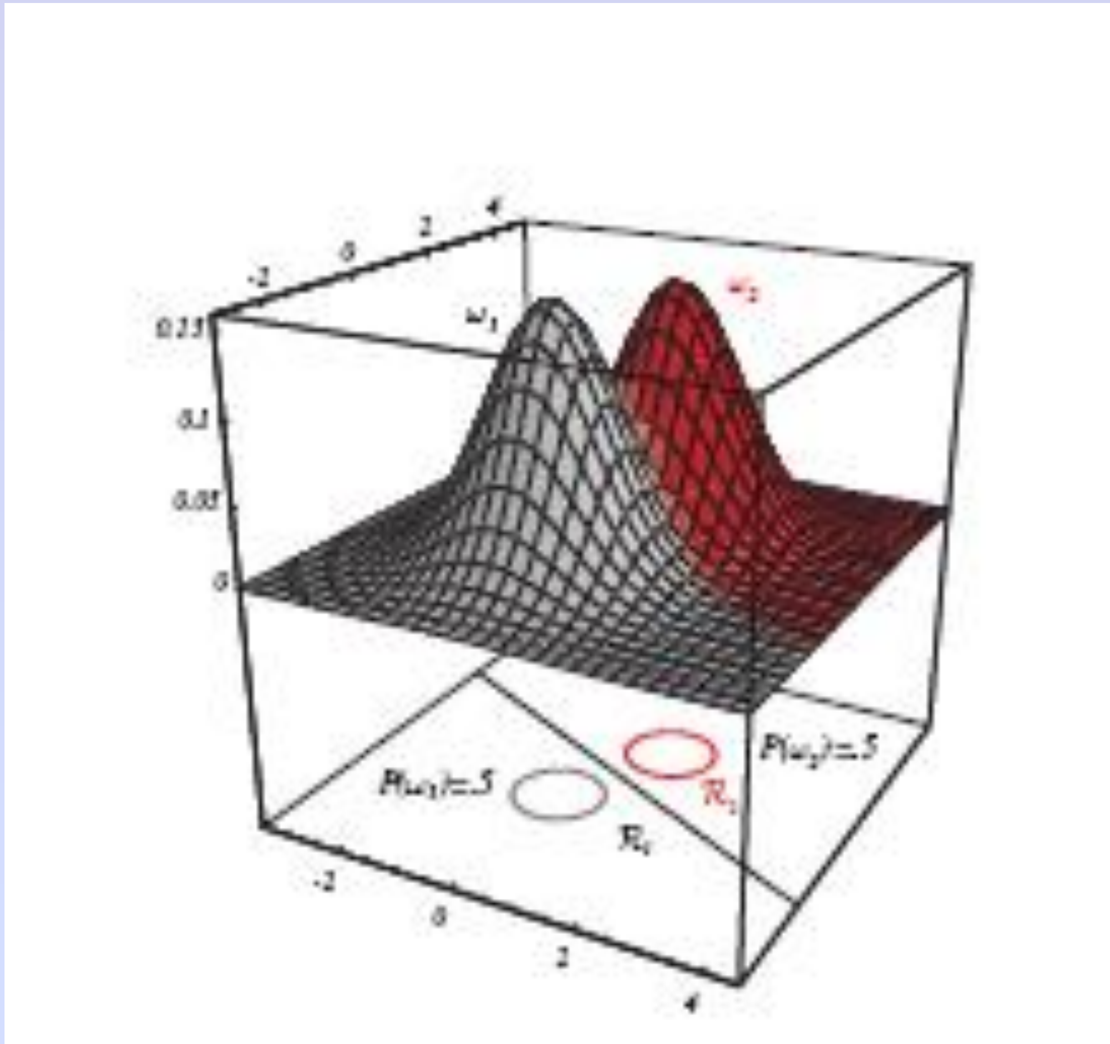
$$\underline{x} \qquad - \; feature\ vector\ (random\ vector,\ d\ \ components)$$

$$P(\underline{x}\,|\,\omega_j) \quad - conditional\ \ probability\ of\ \ \underline{x}$$

$$P(\omega_j) \; - prior\ \ probalitiy$$

$$P(\omega_j\,|\,\underline{x}) = \frac{P(\underline{x}\,|\,\omega_j)\cdot P(\omega_j)}{\sum_{j=1}^{C} P(\underline{x}\,|\,\omega_j)\cdot P(\omega_j)}$$

דוגמא



תורת ההחלטות הביאסינית

הסיכון הכולל

R = Sum of all $R(\alpha_i \mid x)$ for i = 1,…,a

Minimizing R $\Longleftrightarrow$ Minimizing $R(\alpha_i \mid x)$ for i = 1,…, a

$$R(\alpha_i \mid x) = \sum_{j=1}^{j=c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

for i = 1,…,a

תורת ההחלטות הביאסינית

- בכל פעם שניתקל בתצפית מסויימת x נוכל למזער את העלות הצפוייה על-ידי בחירת הפעולה α_i ועל-ידי כך נמזער את הסיכון המותנה

- נראה עתה שפרוצדורת החלטת Bayes היא בעלת ביצוע אופטימלי
- כלל ההחלטה הוא פונקציה $\underline{x}$ α האומרת איזה פעולה יש לבצע לכל וקטור תכונות $\underline{x}$ נתון.
- באופן ספיציפי : איזה פעולה לבצע מתוך הפעולות האפשריות

- הסיכון הכולל – הוא העלות הצפוייה הקשורה לכלל החלטה מסויים

$$R=\int R\ \alpha\ \underline{x}\ |\underline{x}\ p\ \underline{x}\ dx$$

$$dx - d - \textbf{dimensional volume unit}$$

תורת ההחלטות הביאסינית

ברור שאם נבחר את $\alpha \ \underline{x}$ כך ש $R \ \alpha_i | \underline{x}$ יהיה מינימלי,
נמזער את הסיכון הכולל.

זה מצדיק את הכלל הבא, המשיג את הביצוע המיטבי

$$R \ \alpha_i | \underline{x} \ = \sum_{j=1}^{C} \lambda \ \alpha_i | \omega_j \ P \ \omega_j | \underline{x}$$

$$i = 1, 2, \dots, a$$

$$R^* = \min \ R \ \alpha_i | \underline{x}$$

תורת ההחלטות הביאסינית

- בחרו בפעולה α_i עבורה $R(\alpha_i \mid x)$ הוא מינימלי
- R^* הוא מינימלי ונקרא במקרה זה Bayes risk
- זהו הביצוע הטוב ביותר שאפשר להשיג

תורת ההחלטות הביאסיאנית

- סווג לשתי קטיגוריות

- α_1 : deciding ω_1
- α_2 : deciding ω_2

משיקולי פשטות נסמן: $\lambda(\alpha_i \mid \omega_j) = \lambda_{ij}$, העלות של נקיטת הפעולה i כשמצב הטבע הוא j

- העלות היא בהחלטה ω_i כאשר מצב הטבע הוא ω_j

- Conditional risk:
- $R(\alpha_1 \mid x) = \lambda_{11}P(\omega_1 \mid x) + \lambda_{12}P(\omega_2 \mid x)$: העלות (הסיכון) המותנית
- $R(\alpha_2 \mid x) = \lambda_{21}P(\omega_1 \mid x) + \lambda_{22}P(\omega_2 \mid x)$

תורת ההחלטות הביאסיאנית

- נשתמש בכלל הבא:

אם $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$

עבור הפעולה α_1: "בוחרים ω_1"

- תוצאה זו שקולה לכלל הבא:

נחליט ω_1 אם:

- $(\lambda_{21}- \lambda_{11})\, P(x \mid \omega_1)\, P(\omega_1) >(\lambda_{12}- \lambda_{22})\, P(x \mid \omega_2)\, P(\omega_2)$

- אחרת נחליט ω_2

תורת ההחלטות הביאסינית

Likelihood ratio:

הפרוצדורה הזאת שקולה לכלל הבא:

$$if \quad \frac{P(x|\omega_1)}{P(x|\omega_2)} > \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

Likelihood ratio Threshold

אם אי השוויון מתקיים:מבצעים α_1 כלומר מחליטים ω_1

אחרת מבצעים α_2 כלומר מחליטים ω_2

כלומר ההחלטה תהיה α_1 (decide ω_1) אם יחס הסבירויות עולה על הסף (אגף ימין, שאינו תלוי במשתנה x) .

תורת ההחלטות הביאסיאנית

דוגמא

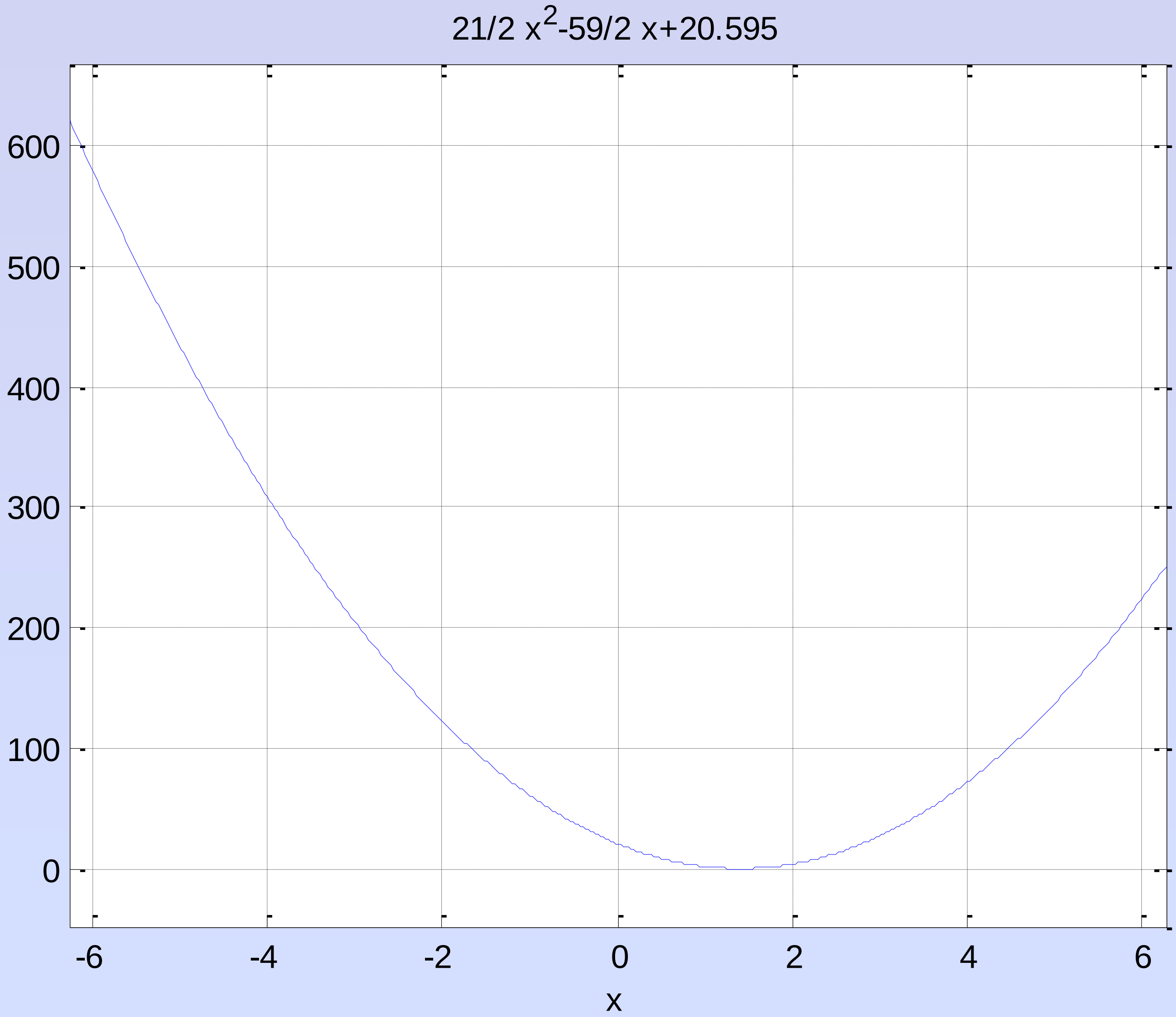
- בחרו בהחלטה האופטימלית עבור

$$\Omega = \ \omega_1, \omega_2$$

- $P(x \mid \omega_1)$ → N(2, 0.5) (Normal distribution)
- $P(x \mid \omega_2)$ → N(1.5, 0.2)
- $P(\omega_1) = 2/3$
- $P(\omega_2) = 1/3$

$$\lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$$

תורת ההחלטות הביאסינית



תורת ההחלטות הביאסינית

סווג קצב שגיאה מינימלי:

Minimum error-rate classification

כל אחד ממצבי הטבע קשור בדרך כלל למחלקות שונות C

אם מבצעים α_i כשמצב הטבע האמיתי הוא ω_j

אז ההחלטה היא נכונה אם i=j ואחרת זוהי שגיאה

אם יש צורך להימנע משגיאות, מחפשים כלל החלטה שימזער את הסתברות השגיאה, כלומר את קצב השגיאה (error rate)

פונקציית העלות או ההפסד במקרה זה מכונה פונקציית עלות סימטרית

$$\lambda \;\; \alpha_i \,|\, \omega_j \;\; = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}, \quad i, j = 1, 2, \ldots C$$

תורת ההחלטות הביאסינית

$$\lambda\ \alpha_i \,|\, \omega_j\ = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}, \quad i, j = 1, 2, \ldots C$$

פונקציית מחיר זו מתאימה מחיר 0 להחלטה נכונה ומחיר של 1 לכל שגיאה, כלומר, לכל השגיאות מחיר זהה:

החלטה נכונה – עלות 0

החלטה לא נכונה – עלות 1

הסיכון הקשור לפונקציית העלות – זוהי ההסתברות הממוצעת לשגיאה, כי הסיכון המותנה הוא:

$$R\ \alpha_i \,|\, \underline{x}\ = \sum_{j=1}^{C} \lambda\ \alpha_i \,|\, \omega_j\ P\ \omega_j \,|\, \underline{x}$$

$$= \sum_{j \neq i} P\ \omega_j \,|\, \underline{x}\ = 1 - P\ \omega_i \,|\, \underline{x}$$

נשים לב ש- P באגף שמאל היא ההסתברות המותנית שהפעולה ה- i היא הנכונה. כלומר הסיכון הוא שהפעולה אינה נכונה.

תורת ההחלטות הביאסינית

$$R\ \alpha_i \,|\, \underline{x}\ = \sum_{j=1}^{C} \lambda\ \alpha_i \,|\, \omega_j\ P\ \omega_j \,|\, \underline{x}$$

$$= \sum_{j \neq i} P\ \omega_j \,|\, \underline{x}\ = 1 - P\ \omega_i \,|\, \underline{x}$$

- כלל ההחלטה הביאסיאני למיזעור הסיכון המותנה דורש לבחור את הפעולה שתמזער את הסיכון המותנה
- למיזעור הסתברות השגיאה הממוצעת צריך לבחור את ה- i שיבצע מקסימיזציה של $P\ \omega_i \,|\, \underline{x}$ - ההסתברות האפוסטריורית

לפיכך לקצב שגיאה מינימלי:

- Decide ω_i *if* $P\ (\omega_i \,|\, x) > P(\omega_j \,|\, x)\ \ \forall j \neq i$