

یادگیری پیازی

قوانین تدریسی (وابستگی)

مقدمه‌ای بر
یادگیری ماشین
بخش سوم

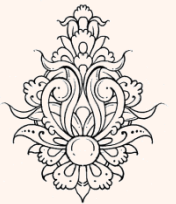


دانشگاه شهید بهشتی
پژوهشکده‌ی فضای مجازی
پاییز ۱۳۹۲
احمد محمودی ازناوه

فهرست مطالب

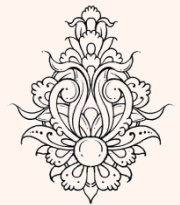
• یادگیری بیزی

- معیارهای تصمیم‌گیری
- قوانین تداعی (وابستگی)



احتمال و استنتاج

- داده‌هایی که مورد استفاده قرار می‌دهیم، حاصل فرآیندی است که (کاملاً) شناخته شده نیست.
- در پدیده‌های تصادفی، متغیرهای غیرقابل مشاهده، موجب پیدایش عدم قطعیت می‌شود.
- $x=f(z)$
- با توجه به این که چنین فرآیندهایی بدین شیوه قابل مدل کردن نیستند، فروجی را به صورت یک متغیر تصادفی تعریف می‌کنیم:
- $P(X=x)$
- بر اساس نمونه‌های ورودی می‌توان این توزیع را تخمین زد، به عنوان مثال برای سکه



$$p_o = \# \{Heads\} / \# \{Tosses\} = \sum_t x^t / N$$

• مسأله‌ی دسته‌بندی اعتبار مشتریان:

– ورودی: درآمد و پس‌انداز

– خروجی: مشتری High risk و low risk

– Input: $\mathbf{x} = [x_1, x_2]^T$, Output: $C \in \{0, 1\}$

– پیش‌بینی:

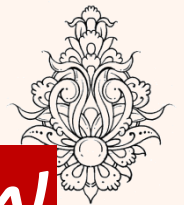
– high risk ($C=1$) or low risk ($C=0$)

choose $\begin{cases} C = 1 \text{ if } P(C = 1 | x_1, x_2) > 0.5 \\ C = 0 \text{ otherwise} \end{cases}$

or

choose $\begin{cases} C = 1 \text{ if } P(C = 1 | x_1, x_2) > P(C = 0 | x_1, x_2) \\ C = 0 \text{ otherwise} \end{cases}$

امتثال شرطی



دسته‌بندی (ادامه...)

- با فرض این ورودی x ، متخیر مشاهده شده است، مسأله یافتن احتمال $P(C|x)$ است.

Bayes' Rule

posterior

احتمال پسین

با چه احتمالی C ، کلاس مربوط به x است.

احتمال پیشین

prior

درست‌نمایی کلاس

Class likelihood

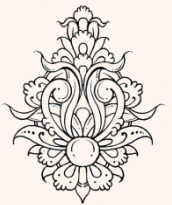
با چه احتمالی x توسط کلاس C تولید می‌شود.

$$P(C|x) = \frac{P(C)p(x|C)}{p(x)}$$

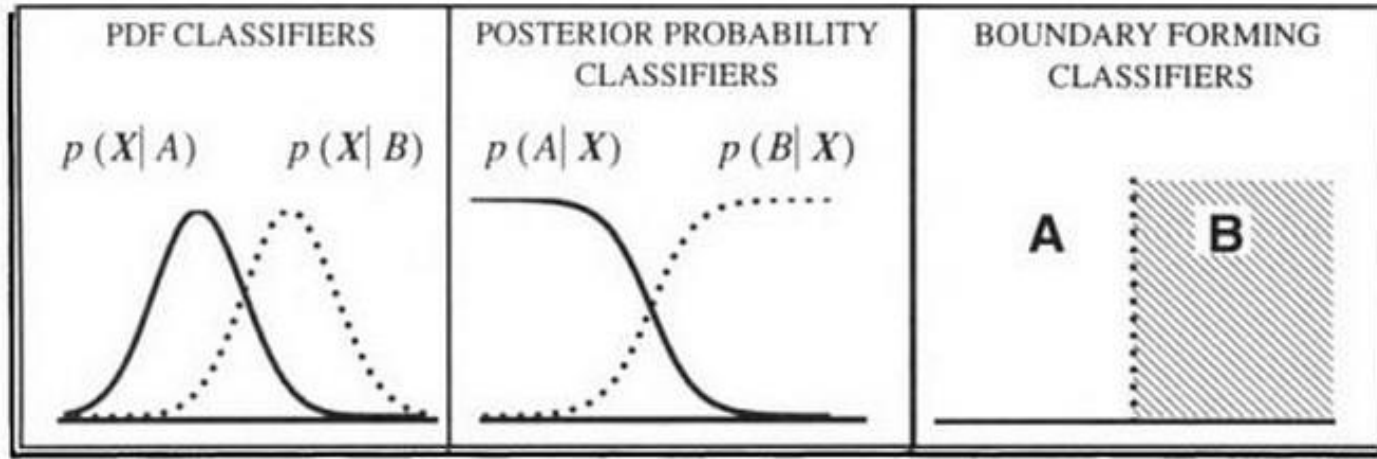
evidence

$$P(C=0) + P(C=1) = 1$$

$$p(x) = p(x|C=1)P(C=1) + p(x|C=0)P(C=0)$$

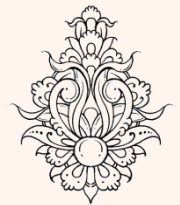


دسته بندی (ادامه...)



pattern recognition using neural networks theory and algorithms for engineers and scientists, by Carl G. Looney

$$P(C | \mathbf{x}) = \frac{P(C) p(\mathbf{x} | C)}{p(\mathbf{x})}$$



دسته بندی چندکلاسی

امتمال رخداد x هنگامی که می دانیم به کلاس C_i تعلق دارد

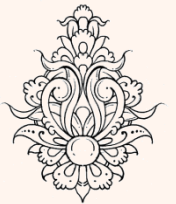
Class likelihood

$$P(C_i | \mathbf{x}) = \frac{P(C_i) p(\mathbf{x} | C_i)}{p(\mathbf{x})}$$

$$P(C_i | \mathbf{x}) = \frac{P(C_i) p(\mathbf{x} | C_i)}{p(\mathbf{x})} = \frac{P(C_i) p(\mathbf{x} | C_i)}{\sum_{k=1}^K P(C_k) p(\mathbf{x} | C_k)}$$

$$P(C_i | \mathbf{x}) = \max_k P(C_k | \mathbf{x})$$

در این صورت کلاس C_i انتخاب می شود.

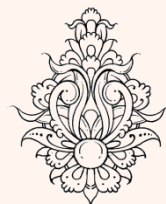


Losses and Risks

- در برخی موارد، تصمیم‌ها پی‌آمد یکسانی ندارند.
– «کنش α_i » به عنوان انتخاب کلاس C_i تعریف شده است.
- λ_{ik} به عنوان میزان ریسک انتخاب کلاس i در زمانی که ورودی به این کلاس k تعلق دارد.
- در این صورت، **expected risk** به صورت زیر محاسبه می‌شود:

$$R(\alpha_i|\mathbf{x}) = \sum_{k=1}^K \lambda_{ik} P(C_k|\mathbf{x})$$

$$\text{choose } \alpha_i \text{ if } R(\alpha_i|\mathbf{x}) = \min_k R(\alpha_k|\mathbf{x})$$

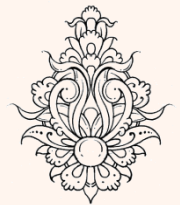


بررسی 0/1 Loss

$$\lambda_{ik} = \begin{cases} 0 & \text{if } i = k \\ 1 & \text{if } i \neq k \end{cases}$$

$$\begin{aligned} R(\alpha_i | \mathbf{x}) &= \sum_{k=1}^K \lambda_{ik} P(C_k | \mathbf{x}) \\ &= \sum_{k \neq i} P(C_k | \mathbf{x}) \\ &= 1 - P(C_i | \mathbf{x}) \end{aligned}$$

برای داشتن کم‌ترین ریسک **محتمل‌ترین** حالت را انتخاب می‌کنیم



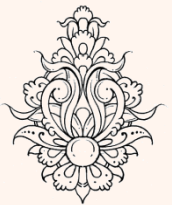
هزینه‌ی بالای انتخاب اشتباه

- در برخی کاربردها، انتخاب اشتباه کلاس هزینه‌ی بالایی دارد، به نحوی که بهتر است هیچ انتخابی توسط سیستم خودکار صورت نپذیرد. در این حالت نمونه به عنوان «مشکوک» تلقی شده و «رد» می‌شود.

– «کنش» جدیدی تعریف می‌شود: رد (reject) : α_{k+1}

choose C_i if $R(\alpha_i|\mathbf{x}) < R(\alpha_k|\mathbf{x}) \quad \forall k \neq i$ and
 $R(\alpha_i|\mathbf{x}) < R(\alpha_{k+1}|\mathbf{x})$

reject $R(\alpha_{k+1}|\mathbf{x}) < R(\alpha_i|\mathbf{x}) \quad i = 1, 2, \dots, k$



هزینه‌ی بالای انتخاب اشتباه (ادامه...)

- به عنوان مثال تابع ریسک به صورت زیر تعریف می‌شود:

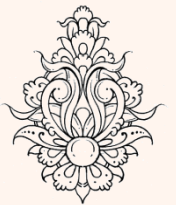
$$\lambda_{ik} = \begin{cases} 0 & \text{if } i = k \\ \lambda & \text{if } i = K + 1, \quad 0 < \lambda < 1 \\ 1 & \text{otherwise} \end{cases}$$

$$R(\alpha_{K+1} | \mathbf{x}) = \sum_{k=1}^K \lambda P(C_k | \mathbf{x}) = \lambda$$

$$R(\alpha_i | \mathbf{x}) = \sum_{k \neq i} P(C_k | \mathbf{x}) = 1 - P(C_i | \mathbf{x})$$

choose C_i if $P(C_i | \mathbf{x}) > P(C_k | \mathbf{x}) \quad \forall k \neq i$ and $P(C_i | \mathbf{x}) > 1 - \lambda$

reject otherwise



Discriminant Functions

choose C_i if $g_i(\mathbf{x}) = \max_k g_k(\mathbf{x})$

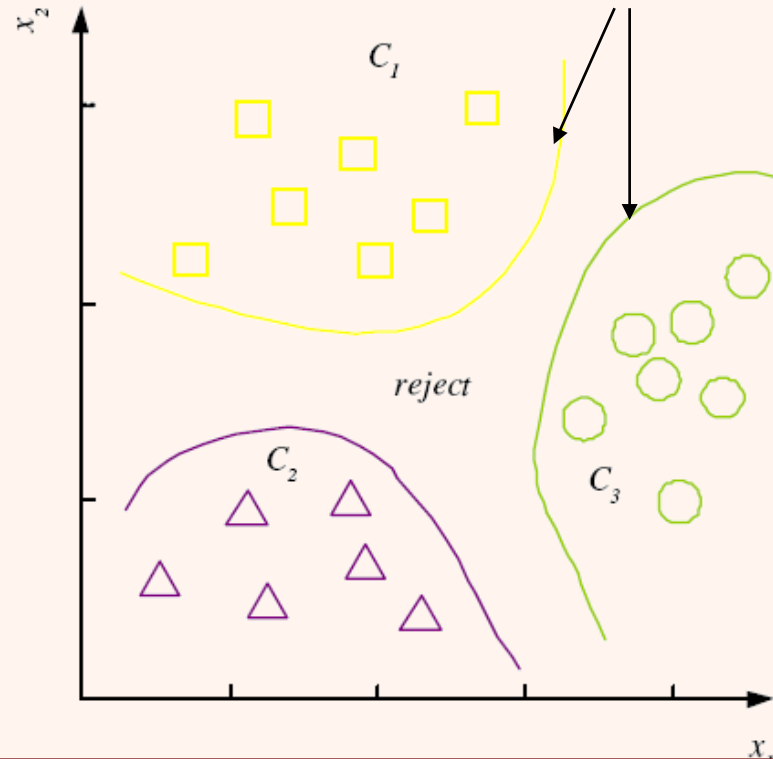
$$g_i(\mathbf{x}) = \begin{cases} -R(\alpha_i | \mathbf{x}) \\ P(C_i | \mathbf{x}) \\ p(\mathbf{x} | C_i)P(C_i) \end{cases}$$

K decision regions $\mathcal{R}_1, \dots, \mathcal{R}_K$

$$\mathcal{R}_i = \{\mathbf{x} | g_i(\mathbf{x}) = \max_k g_k(\mathbf{x})\}$$

توابع جداساز

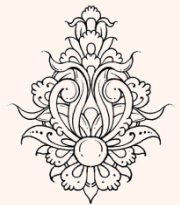
$g_i(\mathbf{x}), i = 1, \dots, K$



Dichotomizer

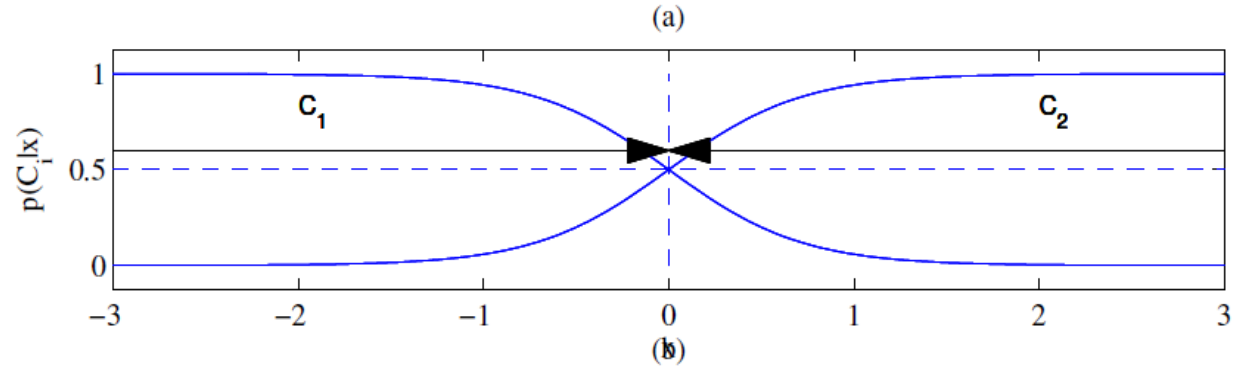
$$g(x) = g_1(x) - g_2(x)$$

choose $\begin{cases} C_1 & \text{if } g(\mathbf{x}) > 0 \\ C_2 & \text{otherwise} \end{cases}$

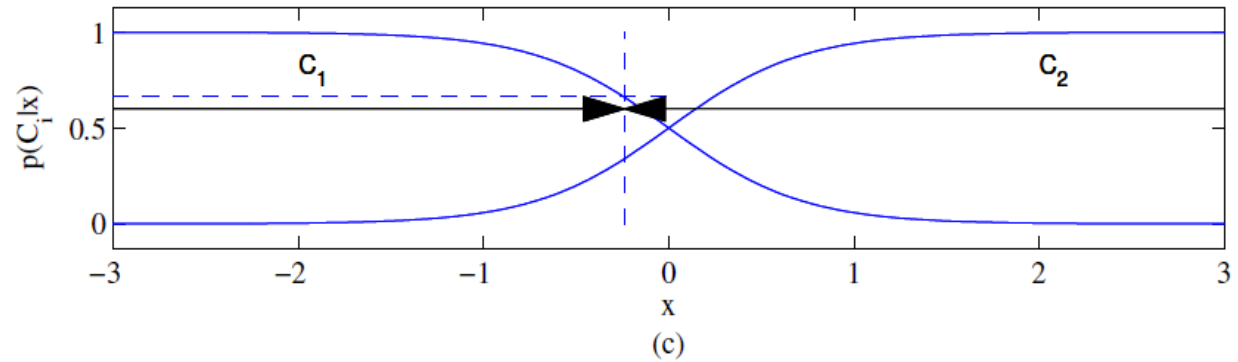


بررسی حالات مختلف

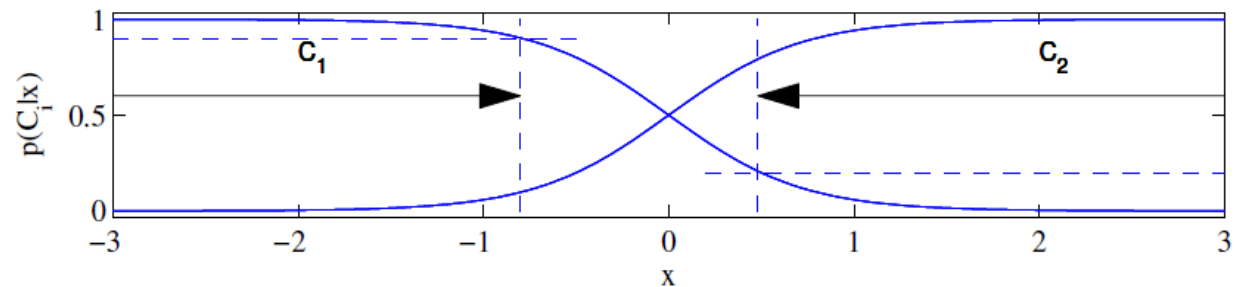
Equal losses



Unequal losses



With reject



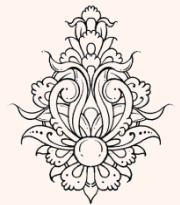
- احتمال این که با در اختیار داشتن ورودی x ، در حالت S_k باشیم: $P(S_k | x)$
- سودمندی کنش α_i وقتی در حالت k ام هستیم:

$$- U_{ik}$$

$$EU(\alpha_i | \mathbf{x}) = \sum_k U_{ik} P(S_k | \mathbf{x})$$

Choose α_i if $EU(\alpha_i | \mathbf{x}) = \max_j EU(\alpha_j | \mathbf{x})$

Expected utility



- «قانون تداعی» (قانون وابستگی) الگوهایی را که بر اساس آن یک رویداد به دیگری مربوط می‌شود، جستجو می‌کند، به عنوان مثال وابستگی خرید قلم به خرید کاغذ.

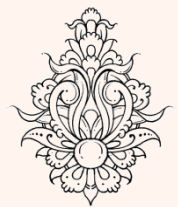
– یک روش مناسب برای یافتن روابط بین متغیرهای موجود در مجموعه داده‌های بزرگ است.

Association rule: $X \rightarrow Y$

مقدمه

antecedent

consequent تالی

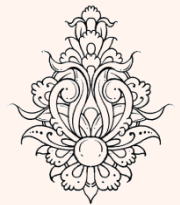


معیارهای وابستگی - پشتیبان

- Support ($X \rightarrow Y$):

$$P(X, Y) = \frac{\#\{\text{customers who bought } X \text{ and } Y\}}{\#\{\text{customers}\}}$$

- در صورتی وابستگی $X \rightarrow Y$ اهمیت خواهد داشت، که نسبت تراکنش‌های X و Y به مجموعه‌ی کل تراکنش‌ها مقدار قابل قبولی باشد.
- معیار «پشتیبان» اهمیت آماری قانون مورد نظر را نشان می‌دهد.



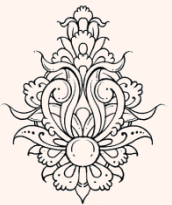
معیارهای وابستگی- اطمینان

- Confidence ($X \rightarrow Y$):

$$P(Y | X) = \frac{P(X, Y)}{P(X)}$$

$$= \frac{\#\{\text{customers who bought } X \text{ and } Y\}}{\#\{\text{customers who bought } X\}}$$

- این معیار طبیعی‌ترین چیزی است که محاسبه می‌شود، در واقع بیانگر این است که تا چه مدی قانون قابل اطمینان است. به بیان دیگر «قدرت» قانون را نشان می‌دهد.
- برای «اطمینان» کافی، باید این معیار به یک نزدیک بوده و مقدار آن تا حد قابل قبولی از $P(Y)$ بیشتر باشد.



معیارهای وابستگی - lift

- Lift/interest ($X \rightarrow Y$):

$$= \frac{P(X, Y)}{P(X)P(Y)} = \frac{P(Y | X)}{P(Y)}$$

– Lift=1

– X و Y مستقل هستند.

– Lift>1

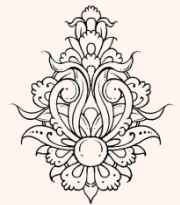
– X موجب افزایش رخداد Y می‌شود.

– Lift<1

– X احتمال بروز Y را کاهش می‌دهد.

این روابط به راحتی قابل تعمیم به بیش از دو آیتم (itemset) است.

$$(X, Z \rightarrow Y): P(Y|X, Z)$$



مثال

Transaction	Items in basket
1	milk, bananas, chocolate
2	milk, chocolate
3	milk, bananas
4	chocolate
5	chocolate
6	milk, chocolate

SOLUTION:

milk → bananas : Support = 2/6, Confidence = 2/4

bananas → milk : Support = 2/6, Confidence = 2/2

milk → chocolate : Support = 3/6, Confidence = 3/4

chocolate → milk : Support = 3/6, Confidence = 3/5



- این الگوریتم در دو مرحله انجام می‌پذیرد:
 - یافتن itemset های مکرر (frequent) (آن‌هایی که معیار پشتیبان بالایی دارند).
 - در صورتی که (X, Y, Z) دارای میزان پشتیبان بالایی باشند، (X, Y) ، (X, Z) و (Y, Z) هم باید دارای پشتیبان بالایی باشند.

Anti-monotone property

- اگر یک itemset، مکرر نباشد، هیچ‌کدام از supersets آن مکرر نخواهند بود.
- یافتن قانون وابستگی بین itemset های یافت شده
 - تقسیم آیتم‌ها به دو دسته‌ی تالی و مقدمه

