Machine Learning

یادگیری بیزی

قوانین تداعی (وابستگی)

یادگیری ماشین
(۱۳-۱۱-۸۰۵-۰۱)
بخش سوم

دانشگاه شهید بهشتی
پژوهشکده‌ی فضای مجازی
پاییز ۱۳۹۶
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 (کاملا) شناخته شده نیست.
- در پدیده‌های تصادفی، متغیرهای غیرقابل مشاهده، موجب پیدایش عدم قطعیت می‌شود.
- $x=f(\mathbf{z})$
- با توجه به این که چنین فرآیندهایی بدین‌شیوه قابل مدل کردن نیستند، خروجی 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

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

دسته‌بندی

- مسأله‌ی دسته‌بندی اعتبار مشتریان:
  - ورودی: درآمد و پس‌انداز
  - خروجی: مشتری low risk و High risk
  - Input: $\boldsymbol{x} = [x_1, x_2]^T$,Output: $C \in \{0,1\}$
  - پیش‌بینی:
  - high risk(C=1) or low risk(C=0)

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

احتمال شرطی

or

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

دسته‌بندی (ادامه...)

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

Bayes’ Rule

درست‌نمایی کلاس

Class likelihood

با چه احتمالی x توسط کلاس C تولید می‌شود.

احتمال پیشین

prior

posterior

احتمال پسین

با چه احتمالی C، کلاس مربوط به x است.

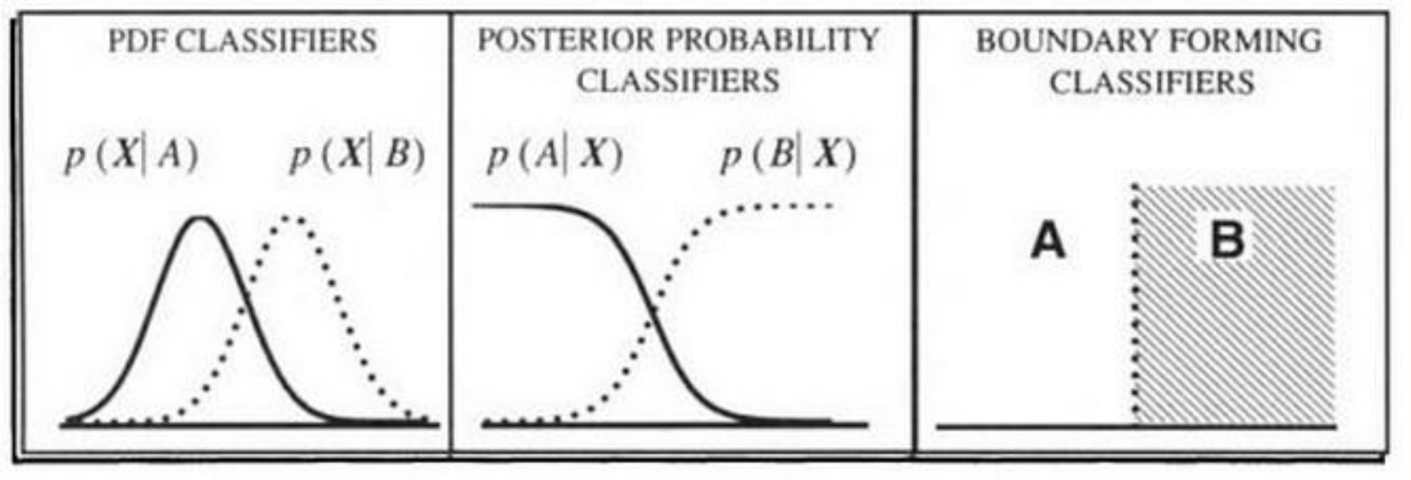
$$\underbrace{P(C|\mathbf{x})}_{\text{posterior}}=\frac{\overbrace{P(C)}^{\text{prior}}\,\overbrace{p(\mathbf{x}|C)}^{\text{Class likelihood}}}{\underbrace{p(\mathbf{x})}_{\text{evidence}}}$$

evidence

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

$$p(\mathbf{x})=p(\mathbf{x}|C=1)P(C=1)+p(\mathbf{x}|C=0)P(C=0)$$

دسته‌بندی (ادامه...)



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

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

دسته‌بندی چندکلاسی

احتمال رخداد x هنگامی که می‌دانیم به کلاس C_i تعلق دارد
Class likelihood

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

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

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

در این صورت کلاس C_i انتخاب می‌شود.

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:

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

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

بررسی 0/1 Loss

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

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

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}

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

$$\text{reject} \qquad R(\alpha_{k+1}|\boldsymbol{x}) < R(\alpha_i|\boldsymbol{x}) \quad i = 1,2,\ldots,k$$

هزینه‌ی بالای انتخاب اشتباه (ادامه...)

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

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

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

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

$$\begin{array}{ll} \text{choose } C_i & \text{if } P(C_i | \mathbf{x}) > P(C_k | \mathbf{x}) \ \forall k \neq i \text{ and } P(C_i | \mathbf{x}) > 1 - \lambda \\ \text{reject} & \text{otherwise} \end{array}$$

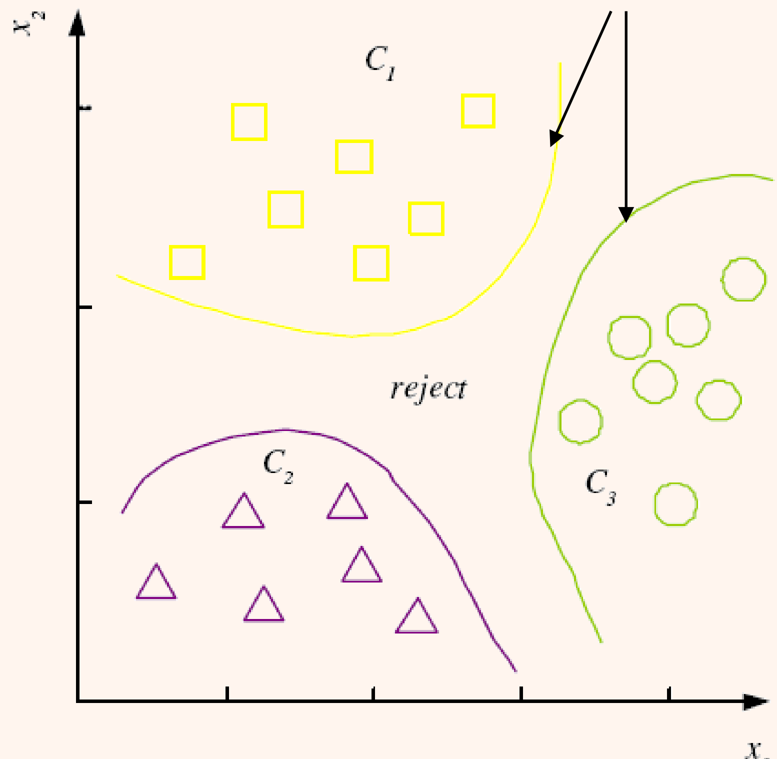
Discriminant Functions

توابع جداساز

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

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

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



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

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

Dichotomizer

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

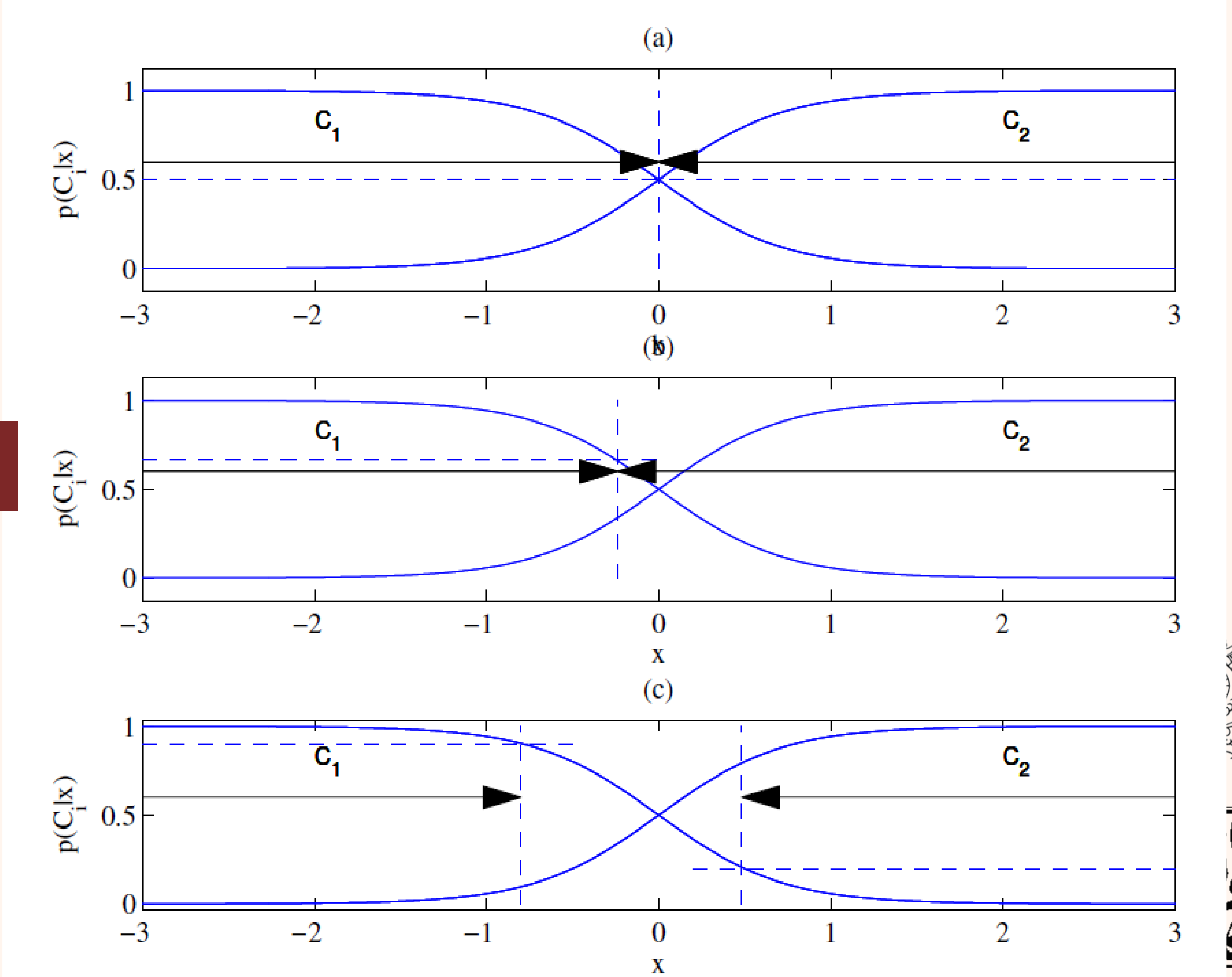
$$\text{choose} \begin{cases} C_1 & \text{if } g(\mathbf{x}) > 0 \\ C_2 & \text{otherwise} \end{cases}$$

بررسی حالات مختلف

Equal losses

Unequal losses

With reject



نظریه‌ی سودمندی

- احتمال این که با در اختیار داشتن ورودی x، در حالت S_k باشیم: $P\left(S_k | x\right)$
- سودمندی کنش α_i وقتی در حالت ikام هستیم:
  - U_{ik}

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

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

Expected utility

قوانین تداعی

Association Rules

- «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 مجموعه‌داده‌های بزرگ است.

Association rule: $X \rightarrow Y$

مقدم antecedent

تالی consequent

معیارهای وابستگی- پشتیبان

- Support $(X \rightarrow Y)$:

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

- در صورتی وابستگی $Y \rightarrow X$ 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 ($X \to Y$):

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

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

- این معیار طبیعی‌ترین چیزی است که محاسبه می‌شود، در واقع بیانگر این است که تا چه حدی 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 ($X \rightarrow Y$):

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

  - 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:

$$\begin{aligned}
\text{milk} \rightarrow \text{bananas} &: \text{Support} = 2/6, \text{Confidence} = 2/4 \\
\text{bananas} \rightarrow \text{milk} &: \text{Support} = 2/6, \text{Confidence} = 2/2 \\
\text{milk} \rightarrow \text{chocolate} &: \text{Support} = 3/6, \text{Confidence} = 3/4 \\
\text{chocolate} \rightarrow \text{milk} &: \text{Support} = 3/6, \text{Confidence} = 3/5
\end{aligned}$$

Apriori algorithm (Agrawal et al., 1996)

• این الگوریتم در دو مرحله انجام می‌پذیرد:

۱

– یافتن itemsetهای 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 supersetهای آن مکرر نخواهند بود.

۲

– یافتن قانون وابستگی بین itemsetهای یافت شده

  • تقسیم آیتم‌ها به دو دسته‌ی تالی و مقدم