Matematické okienko

(Strojové učenie II)

M. Mach

Katedra kybernetiky a umelej inteligencie, FEI, TUKE

január-február 2023

Stredná hodnota (1)

- $E[X]$ - stredná hodnota náhodnej veličiny X
  - označuje sa aj ako očakávaná (expected) hodnota
  - náhodná veličina X má pravdepodobnostnú distribúciu $p(x)$
  - niekedy namiesto $E[X]$ sa zapisuje $E_p[X]$
- pre spojitú náhodnú veličinu je $E[X] = \int_{-\infty}^{\infty} x\ p(x) dx$
- pre diskrétnu náhodnú veličinu je pravdepodobnostne váženým priemerom $E[X] = p(x_1)x_1 + \ldots + p(x_n)x_n = \sum_{i=1}^{n} p(x_i)\, x_i$
  - pri rovnakých pravdepodobnostiach degraduje na aritmetický priemer

Stredná hodnota (2)

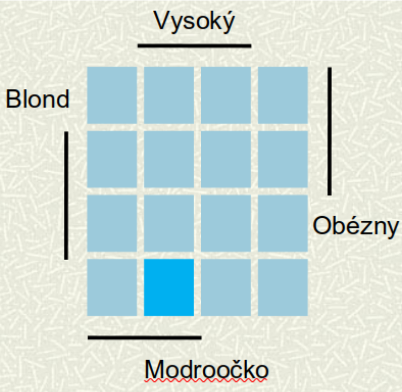
- Vybrané vlastnosti strednej hodnoty
  - nech X, Y a Z sú náhodné premenné, a a b sú konštanty
    $E[a] = a$
    $E[aX] = aE[X]$
    $E[X + Y] = E[X] + E[Y]$
    $E[aX + bY|Z] = aE[X|Z] + bE[Y|Z]$
    $E[E[X]] = E[X]$
  - (the Law of iterated expectations)
    $E[E[X|Y]] = E[X]$
    $E[E[X|Y]\,|Z] = E[X|Z]$

Stredná hodnota (3)

- Dôkaz pre Law of iterated expectations
  - $E[E[X|Y]]$
    $= \sum_y E[X|Y=y]\; p(Y=y)$
    $= \sum_y \sum_x x\, p(X=x|Y=y)\; p(Y=y)$
    $= \sum_y \sum_x x\, p(X=x, Y=y)$
    $= \sum_y \sum_x x\, p(Y=y|X=x)\; p(X=x)$
    $= \sum_y \sum_x x\, p(X=x)\; p(Y=y|X=x)$
    $= \sum_x x\, p(X=x) \sum_y p(Y=y|X=x)$
    $= \sum_x x\, p(X=x) \qquad$ (lebo $\sum_y p(Y=y|X=x) = 1$)
    $= E[X]$

Pravdepodobnosť (1)

- Združená pravdepodobnosť

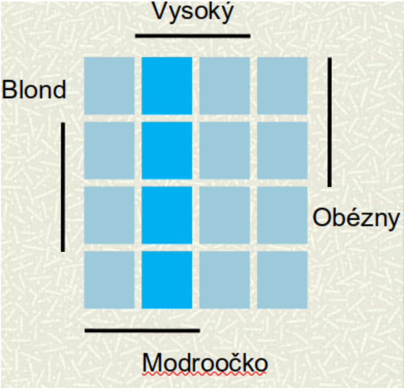


- distribúcia pravdepodobnosti $p(V, B, O, M)$
- pravdepodobnosť $p(V = yes, B = no, O = no, M = yes)$

- $\sum_V \sum_B \sum_O \sum_M p(V, B, O, M) = 1.0$

Pravdepodobnosť (2)

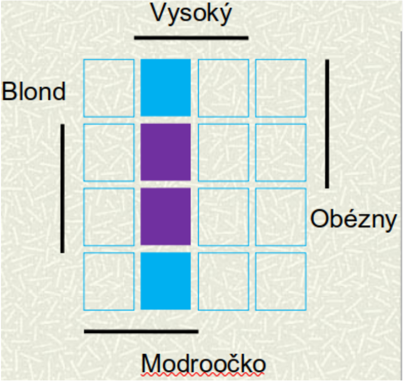
- Marginálna pravdepodobnosť



- distribúcia pravdepodobnosti $p(V, M) = \sum_B \sum_O p(V, B, O, M)$
- pravdepodobnosť $p(V = yes, M = yes) = \sum_B \sum_O p(V = yes, B, O, M = yes)$

- $\sum_V \sum_M p(V, M) = \sum_V \sum_M (\sum_B \sum_O p(V, B, O, M)) = 1.0$

- Podmienená pravdepodobnosť



- distribúcia pravdepodobnosti $p(O, B|V = yes, M = yes)$
- pravdepodobnosť $p(B = yes|V = yes, M = yes)$

- $\sum_V \sum_M p(V, M|B, O) = 1.0$

Pravdepodobnosť (4)

- Vybrané vlastnosti
  - $0.0 \leq p(A = a, B = b) \leq 1.0$
  - $P(A, B) = p(A|B)p(B) = p(B|A)p(B)$
    $P(A, B|Z) = p(A|B, Z)p(B|Z) = p(B|A, Z)p(B|Z)$
  - nezávislosť premenných A a B
    $p(A|B) = p(A)$
    $p(A|B, Z) = p(A|Z)$
    $p(A, B) = p(A)p(B)$
  - veta o úplnej pravdepodobnosti
    $p(A) = \sum_B p(A|B)p(B)$
  - podmienená pravdepodobnosť
    $p(A|C, D) = \frac{p(A,C,D)}{p(C,D)} = \frac{\sum_B p(A,B,C,D)}{\sum_A \sum_B p(A,B,C,D)}$

Markovovské reťazce (1)

- Markovovská vlastnosť (memorylessness): "Budúcnosť pri danej prítomnosti nezáleží na minulosti"

Definition

$$P[S_{t+1}|S_1, \ldots, S_t] = P[S_{t+1}|S_t]$$

- Vlastnosť stochastických procesov, modelujúcich sekvenciu stavov
- Aktuálny stav obsahuje všetku relevantnú informáciu pre predikciu budúceho stavu
  - nezáleží na tom, aká postupnosť stavov viedla k aktuálnemu stavu
  - pri znalosti aktuálneho stavu je možné históriu minulých stavov ignorovať

Markovovské reťazce (2)

- Markovovský proces (reťazec)
  - stochastický model sekvencie možných stavov
  - spĺňa Markovovskú vlastnosť

Markovovský proces je dvojica (S,P), kde

- S je množina stavov
- P je pravdepodobnostná prechodová matica $P_{ss'} = P[S_{t+1} = s' | S_t = s]$

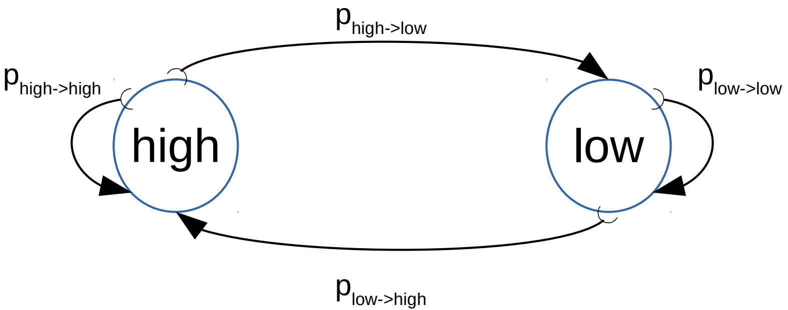
- Prechodová matica definuje pravdepodobnosti prechodov medzi dvojicami stavov

CIT Centrum inteligentných technológií

Markovovské reťazce (3)

- Možná sekvencia: *high*, *high*, *low*, *high*, *low*, ...
- Stavová prechodová matica $P_{ss'}$

$$\begin{array}{cc} & do \\ z & \begin{bmatrix} p_{high->high} & p_{high->low} \\ p_{low->high} & p_{low->low} \end{bmatrix} \end{array}$$



Markovovské reťazce (4)

- Distribúcia stavov v čase t je $\pi(t)$, $\pi(0)$ je počiatočná distribúcia
$$\begin{bmatrix} p_{high} \\ p_{low} \end{bmatrix}$$
- Markovovský reťazec je daný vzťahom
$\pi(n+1) = \pi(n)^T P_{ss'}$
$\pi(n+1) = \pi(0)^T P_{ss'}^{n+1}$
- Absorpčný stav - pravdepodobnosť zotrvania v danom stave je 1
- Komunikujúce stavy - pravdepodobnosť dosiahnutia jedného z druhého (a aj opačným smerom) pomocou sekvencie prechodov je kladná

Markovovské reťazce (5)

- Vlastnosti Markovovských reťazcov
  - (Časovo) Homogénny reťazec - pravdepodobnosti prechodov matice $P_{ss'}$ nezáležia na čase
  - Distribúcia stavov je stacionárna, ak platí
    $\pi = \pi^T P_{ss'}$
  - Regulárny reťazec - distribúcia stavov limituje k stacionárnej distribúcii
    $\lim_{n->\infty} \pi(n) = \pi$
    a tá nezávisí na počiatočnej distribúcii stavov

Nekonečné rady (1)

- Ak je daná postupnosť $\{a_i\}_{i=1}^{\infty}$, tak súčet členov tejto postupnosti je nekonečný rad

Definition

$a_1 + a_2 + \ldots = \sum_{i=1}^{\infty} a_i$

- Nekonečný rad je konvergentný, ak postupnosť čiastočných súčtov $\{s_i\}_{i=1}^{\infty}$ konverguje

Definition

$s = \lim_{n->\infty} s_n = \lim_{n->\infty} \sum_{i=1}^{n} a_i = \sum_{i=1}^{\infty} a_i$

- Ak rad konverguje, tak limita postupnosti $\{a_i\}_{i=1}^{\infty}$ je rovná nule, inak rad diverguje

Nekonečné rady (2)

- Typy nekonečných radov
  - Harmonický - členy sú prevrátené hodnoty $\{1/i\}_{i=1}^{\infty}$ - rad diverguje
  - Aritmetický - aritmetická postupnosť s diferenciou - rad konverguje iba pre nulovú diferenciu a nulovú prvú hodnotu
  - Geometrický - geometrická postupnosť s kvocientom - rad konveguje pre kvocient z intervalu $(-1, 1)$ alebo pre nulovú prvú hodnotu
    $$\sum_{i=1}^{\infty} a_1 q^{i-1} = \frac{a_1}{1-q}$$

Nekonečné rady (3)

- Súčet konečného radu $s_n = \sum_{i=1}^{n} a_1 q^{i-1}$

  $qs_n - s_n = \sum_{i=1}^{n} a_1 q^{i} - \sum_{i=1}^{n} a_1 q^{i-1} = a_1 q^n - a_1$

  $=>$

  $s_n = a_1 \frac{q^n - 1}{q-1}$

- Dôkaz konvergencie nekonečného geometrického radu pre $q \in (-1, 1)$

  $\sum_{i=1}^{\infty} a_1 q^{i-1} =$

  $\lim_{n->\infty} \sum_{i=1}^{n} a_1 q^{i-1} =$

  $\lim_{n->\infty} \left(a_1 \frac{q^n - 1}{q-1} \right) =$

  $a_1 \frac{1}{q-1} \lim_{n->\infty} (q^n - 1) =$

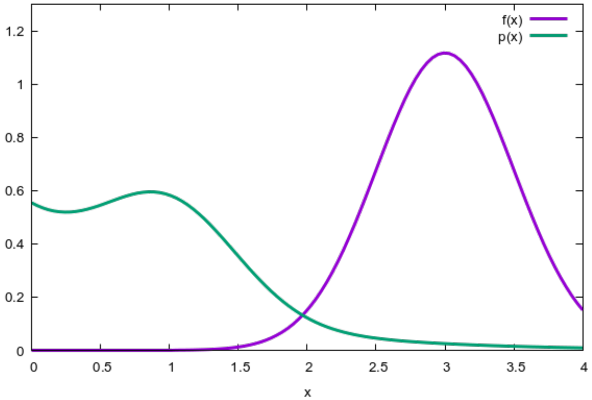
  $a_1 \frac{1}{q-1} (\lim_{n->\infty} q^n - \lim_{n->\infty} 1) =$

  $a_1 \frac{1}{q-1} (0 - 1) =$

  $\frac{a_1}{1-q}$

Vzorkovanie podľa dôležitosti (1)

- Je zadaná funkcia $f(x)$, pričom premenná x je náhodnou premennou s pravdepodobnostnou distribúciou $p(x)$



Vzorkovanie podľa dôležitosti (2)

- Chceme určiť strednú hodnotu $E_p[f(x)]$
- Aproximácia vzorkovaním a priemerom

$$E_p[f(x)] = \int f(x)p(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i) \quad x_i \sim p(x)$$

- Priemer bude náhodnou hodnotou centrovanou okolo $E_p[f(x)]$
- Veľká variancia - hodnoty generované podľa $p(x)$ sú väčšinou mimo "dôležitého" regiónu funkcie $f(x)$
- Zmenšenie variancie
  - použitie veľkého počtu vzoriek
  - väčšia pozornosť "dôležitému" regiónu $f(x)$

Vzorkovanie podľa dôležitosti (3)

- Zavedie sa pravdepodobnostná distribúcia $q(x)$
  - $q(x) > 0$ vždy keď $f(x)p(x) \neq 0$
  - $p(x)/q(x)$ je vzorkovací pomer/váha

$$E_p[f(x)] = \int f(x)p(x)dx = \int \left[f(x)\frac{p(x)}{q(x)}\right] q(x)dx$$

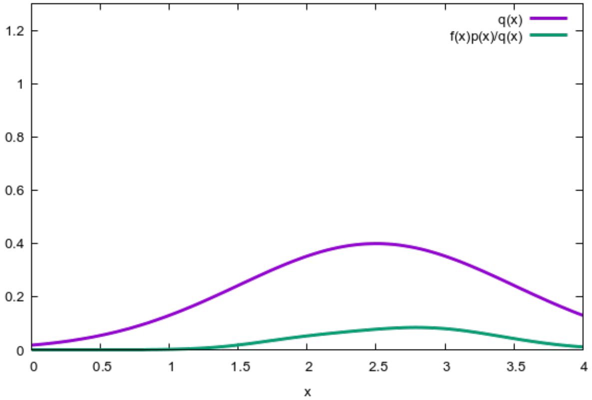
- Aproximácia vzorkovaním a priemerom

$$E_p[f(x)] = E_q\left[f(x)\frac{p(x)}{q(x)}\right] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)} \quad x_i \sim q(x)$$

- Voľba $q(x)$ znižujúca varianciu
  - $q(x)$ veľké všade tam, kde $\mid f(x)p(x) \mid$ je veľké
  - nevhodná voľba môže situáciu aj zhoršiť

Vzorkovanie podľa dôležitosti (4)

- Vhodná voľba $q(x)$ - väčšia pozornosť venovaná "dôležitému" regiónu $f(x)$



Vzorkovanie podľa dôležitosti (5)

- Nevhodná voľba $q(x)$ - ešte menšia pozornosť venovaná "dôležitému" regiónu $f(x)$