机器学习与人工智能　　　　Ensemble Models　　　　

1. 集成学习 Ensemble learner

个体学习器 individual learner

弱学习器 weak learner

一般结构：先产生一组 individual learner，再用某种策略将它们结合起来。

2. 结合策略

① 平均法. $H(x) = \frac{1}{T}\sum_{i=1}^{T} h_i(x)$.

加权平均. $H(x) = \sum_{i=1}^{T} w_i h_i(x)$. $w_i \geq 0$. $\sum w_i = 1$.

↳ 一般从训练数据中求得 (e.g. w_i 与误差成反比).

② 投票法.

classifier. a, majority / hard voting. { the majority (mode)

若 tie. random pick

or select based on the ascending order

$h_i^j(x) \in \{0, 1\}$

$j \to$ outcome class

b, weighted avg probability (soft) $h_i^{(j)}(x) \in [0, 1]$.

return the class label as argmax of the sum of predicted probabilities.

③ 学习法 stacking.

初级学习器 → 次级学习器 (meta learner)

Meta learner training set $D' = \{x^{(i)'}, y^{(i)}\}$. $x^{(i)'} = \{h_1(x^{(i)}), \dots h_T(x^{(i)})\}$.

可能会 overfitting → N-fold CV.

$D = D_j + \bar{D}_j$ 　　 from $\bar{D}_j \Rightarrow h_{t,j}$

test　training.

(jth fold)　　 D_j : $z_{it} = h_{t,j}^{(i)}(x^{(i)})$ any i in D_j

$D' = \{(z_i, y_i)\}_{i=1}^{m}$.

机器学习与人工智能 Hidden Markov Model. Oct 2021 P1.

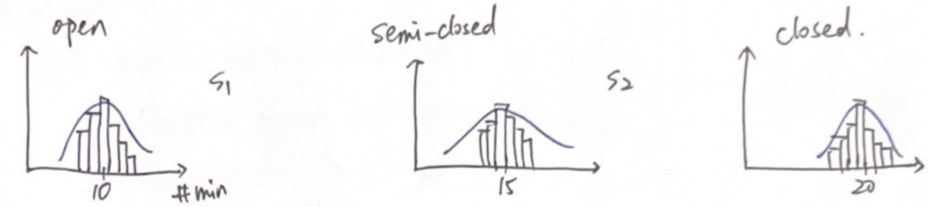
隐马尔可夫模型

Y. 隐状态

X. 观察状态

1. 高架修路. open. semi-closed (一半道封闭). closed.

$x_t \triangleq$ travel time on day t. $y_t \triangleq$ state of the road on day t.



$p(x_t | y_t)$

MLE (multinomial)

$\hat{P}_{open} = \frac{\#(y_t = open)}{365}$ $\hat{P}_{sc} = \frac{\#(y_t = sc)}{365}$ $\hat{P}_{closed} = \frac{\#(y_t = c)}{365}$ $p(y_t)$

$p(x_t, y_t) = p(x_t | y_t) p(y_t)$

2. 1st order Markov Assumption 一阶. 序列数据.

Let y_t = state of system at time t.

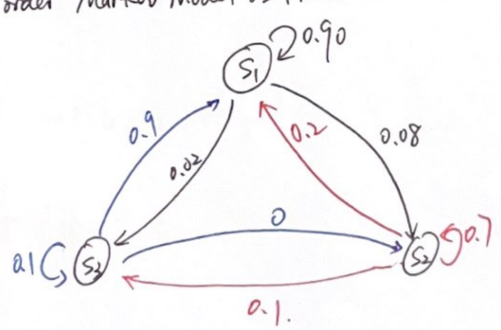
$$P(y_t | y_{t-1}, y_{t-2}, \dots, y_1) = P(y_t | y_{t-1})$$ By Assumption

$$\Rightarrow y_t \perp\!\!\!\perp y_j | y_{t-1}. \ \forall j < t-1.$$

$$\Rightarrow P(y_1, \dots, y_T) = \prod_{t=1}^{T} P(y_t | y_{t-1}, \dots, y_1)$$ by chain Rule 条件概率的链式法则.

$$= \prod_{t=1}^{T} P(y_t | y_{t-1})$$ By 1st order Markov. Assump.

1st order Markov Model as Finite State Machine



$P(y_t | y_{t-1})$.

3. MLE for HMM.

① Def. $D = \{(\vec{x}^{(i)}, \vec{y}^{(i)})\}_{i=1}^{N}$. $\vec{x}^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_T^{(i)}]^{\text{Transpose}}$.

$\vec{y}^{(i)} = [y_1^{(i)}, y_2^{(i)}, \dots, y_T^{(i)}]^{\text{Transpose}}$.

② likelihood.

A. Emission matrix. 条件概率.

B. Transition matrix. 状态转移概率.

C. Initial prob. $P(Y_1 = k) = C_k. \forall k$

$$l(A, B, C) = \sum_{i=1}^{N} \log p(\vec{x}^{(i)}, \vec{y}^{(i)} \mid A, B, C).$$

$$= \sum_{i=1}^{N} \Big[\underbrace{\log p(y_1^{(i)} \mid C)}_{\text{initial}} + \underbrace{\Big(\sum_{t=2}^{T} \log P(y_t^{(i)} \mid y_{t-1}^{(i)}, B)\Big)}_{\text{Transition}} + \underbrace{\Big(\sum_{t=1}^{T} \log P(x_t^{(i)} \mid y_t^{(i)}, A)\Big)}_{\text{emission}} \Big].$$

③ MLE. $\hat{A}, \hat{B}, \hat{C} = \underset{A, B, C}{\arg\max}\ l(A, B, C)$

$$\Rightarrow \hat{C} = \underset{C}{\arg} \sum_{i=1}^{N} \log p(y_1^{(i)} \mid C) \qquad \hat{C}_k = \frac{\#(y_1^{(i)} = k)}{N}.$$

$$\hat{B} = \underset{B}{\arg\max} \sum_{i=1}^{N} \sum_{t=2}^{T} \log P(y_t^{(i)} \mid y_{t-1}^{(i)}, B) \qquad \hat{B}_{jk} = \frac{\#(y_t^{(i)} = k \text{ and } y_{t-1}^{(i)} = j)}{\#(y_{t-1}^{(i)} = j)}$$

$$\hat{A} = \underset{A}{\arg\max} \sum_{i=1}^{N} \sum_{t=1}^{T} \log P(x_t^{(i)} \mid y_t^{(i)}, A) \qquad \hat{A}_{jk} = \frac{\#(x_t^{(i)} = k \ \& \ y_t^{(i)} = j)}{\#(y_t^{(i)} = j)}$$

注 Assume $y_0 = k$, we fold C to B.

EM. 期望最大算法.

B.

4. 3 problems for a HMM.

Evaluation. $p(\vec{x}) = \sum_{\vec{y} \in \mathcal{Y}_{\vec{x}}} p(\vec{x}, \vec{y})$.

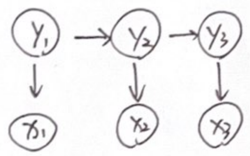
For $|\vec{y}| = T$. and $y_t \in \{1, \dots, k\}$
there are $\underline{\underline{k^T}}$ possible value of $\vec{y}$

(Viterbi) Decoding 解码. $\hat{y} = \arg\max_{\vec{y} \in \mathcal{Y}_{\vec{x}}} p(\vec{y} \mid \vec{x})$.

Learning. $\hat{\lambda} = (A, B)$. 参数估计

↳ MLE. EM (Baum-Welch 算法)

Ex.



healthy. sick. vacation.
$y_t \in \{H, S, V\}$

$x_t \in \{C, N\}$ class or not.

$$p(x_1, x_2, x_3) = \sum_{y_1} \sum_{y_2} \sum_{y_3} p(x_1, x_2, x_3, y_1, y_2, y_3).$$

$$\hat{y}_1, \hat{y}_2, \hat{y}_3 = \arg\max_{y_1, y_2, y_3} p(y_1, y_2, y_3 \mid x_1, x_2, x_3)$$

Joint dist

$p(x_1, x_2, x_3, y_1, y_2, y_3)$

$= p(y_1) p(x_1|y_1) p(x_2|y_2) p(x_3|y_3)$
$p(y_2|y_1) p(y_3|y_2)$

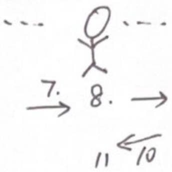
marginal $p(y_2 = V \mid x_1, x_2, x_3) = \sum_{y_1} \sum_{y_3} p(y_1, y_2, y_3 \mid x_1, x_2, x_3)$.

Brute Force for Evaluation 暴力算法.

```
def eval(x).            ← p(x⃗)
    p_x = 0.
    for y in all_y(x):          Y_x.
        p_x += joint(x, y)      p(x⃗, y⃗)
    return p_x
```

5. Forward-Backward Algorithm 前向后向算法.

(1). Motivation 报数.



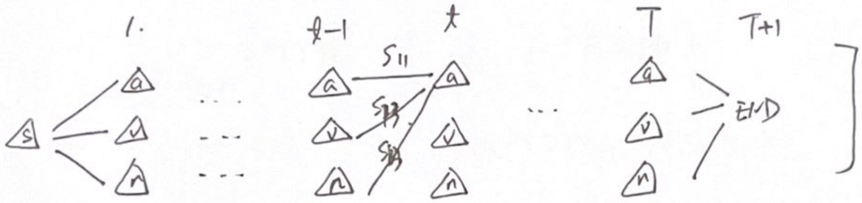
7+10+1 = 18人.

(2). Part-of-speech tagging 词性标注. 对句子中的词语进行分类标注的过程

→NLP的基础任务. (序列标注问题)

应用: 句法分析预处理. 词汇获取/信息~~处理~~抽取 预处理

(3). F-B Algorithm



trelli

forward.

for $t=1,\dots T$.

for $k=1,\dots K$.

$$\alpha_t(k)=\sum_{j=1}^{K}\alpha_{t-1}(j)\,s_{kj}.$$

For HMM

$p(y_t=k \mid y_{t-1}=j)\,p(x_t \mid y_t=k)$.

动态规划 Dynamic Programming.

要解一个给定的问题, 需要解其不同部分(子问题), 再合并子问题的解得到原问题的解.

F-B Algo. $O(K^2T)$

Brute Force $O(K^T)$

$$P(Y_t=k\mid\vec{x}) = \frac{P(x_1,\cdots,x_t\mid Y_t)\,P(x_{t+1},\cdots x_T\mid Y_t)\,P(Y_t)}{P(\vec{x})}$$ 序列字符间相互独立.

$$= P(x_1,\cdots,x_t,Y_t)\,P(x_{t+1},\cdots,x_T\mid Y_t)/P(\vec{x}).$$

Forward Algorithm.

$$\alpha_t(k) = P(x_1,x_2,\cdots,x_t,Y_t=k).$$

$$= P(x_1,\cdots x_t\mid Y_t=k)\,P(Y_t=k).$$

$$= P(x_1,\cdots,x_{t-1}\mid Y_t)\,P(x_t\mid Y_t)\,P(Y_t).$$

$$= P(x_1,\cdots,x_{t-1},Y_t)\,P(x_t\mid Y_t).$$

$$= \sum_{Y_{t-1}} P(x_1,x_2,\cdots,x_{t-1},y_{t-1},Y_t)\,P(x_t\mid y_t).$$

$$= \sum_{Y_{t-1}} P(x_1,\cdots x_{t-1},Y_t\mid Y_{t-1})\,P(Y_{t-1})\,P(x_t\mid Y_t).$$

$$= \sum_{Y_{t-1}} P(x_1,\cdots x_{t-1}\mid Y_{t-1})\,P(Y_t\mid Y_{t-1})\,P(Y_{t-1})\,P(x_t\mid Y_t).$$

$$= \sum_{Y_{t-1}} P(x_1,\cdots,x_{t-1},Y_{t-1})\,P(Y_t\mid Y_{t-1})\,P(x_t\mid Y_t).$$

$$= \sum_{Y_{t-1}} \alpha_{t-1}(\,)\,P(Y_t\mid Y_{t-1})\,P(x_t\mid Y_t).$$

$$= P(x_t\mid Y_t)\sum_{j=1}^{k}\alpha_{t-1}(j)\,P(Y_t\mid Y_{t-1}).$$

Backward Algorithm.

$$\beta_t(k) = P(x_{t+1},x_{t+2},\cdots x_T\mid Y_t).$$

$$= \sum_{j=1}^{K} P(x_{t+1},x_{t+2},\cdots x_T,Y_{t+1}\mid Y_t)$$

$$\cancel{= \sum_{Y_{t+1}} P(x_{t+1},\cdots x_T\mid Y_{t+1})\,P(Y_{t+1}\mid}$$

$$= \sum_{Y_{t+1}} P(x_{t+1},\cdots x_T\mid y_{t+1},y_t)\,P(Y_{t+1}\mid Y_t).$$

$$= \sum_{Y_{t+1}} P(Y_{t+2},\cdots Y_T\mid Y_{t+1})\,P(x_{t+1}\mid Y_{t+1})\,P(Y_{t+1}\mid Y_t).$$

$$= \sum_{j=1}^{k}\beta_{t+1}(j)\,P(X_{t+1}\mid Y_{t+1})\,P(Y_{t+1}\mid Y_t)$$