

阿里云天池牛年读书会

《统计学学习方法》 导读

分享嘉宾：陈亮
七月在线 AI Lab 负责人

天池读书会

TIANCHI 天池



清华大学出版社
TSINGHUA UNIVERSITY PRESS

《统计学习方法》

从ML中的模型关联、符号表及向量化、决策函数与数形结合、概率模型与基本规则四个方面给大家分享本书。

直播嘉宾：陈亮 七月在线AI Lab负责人

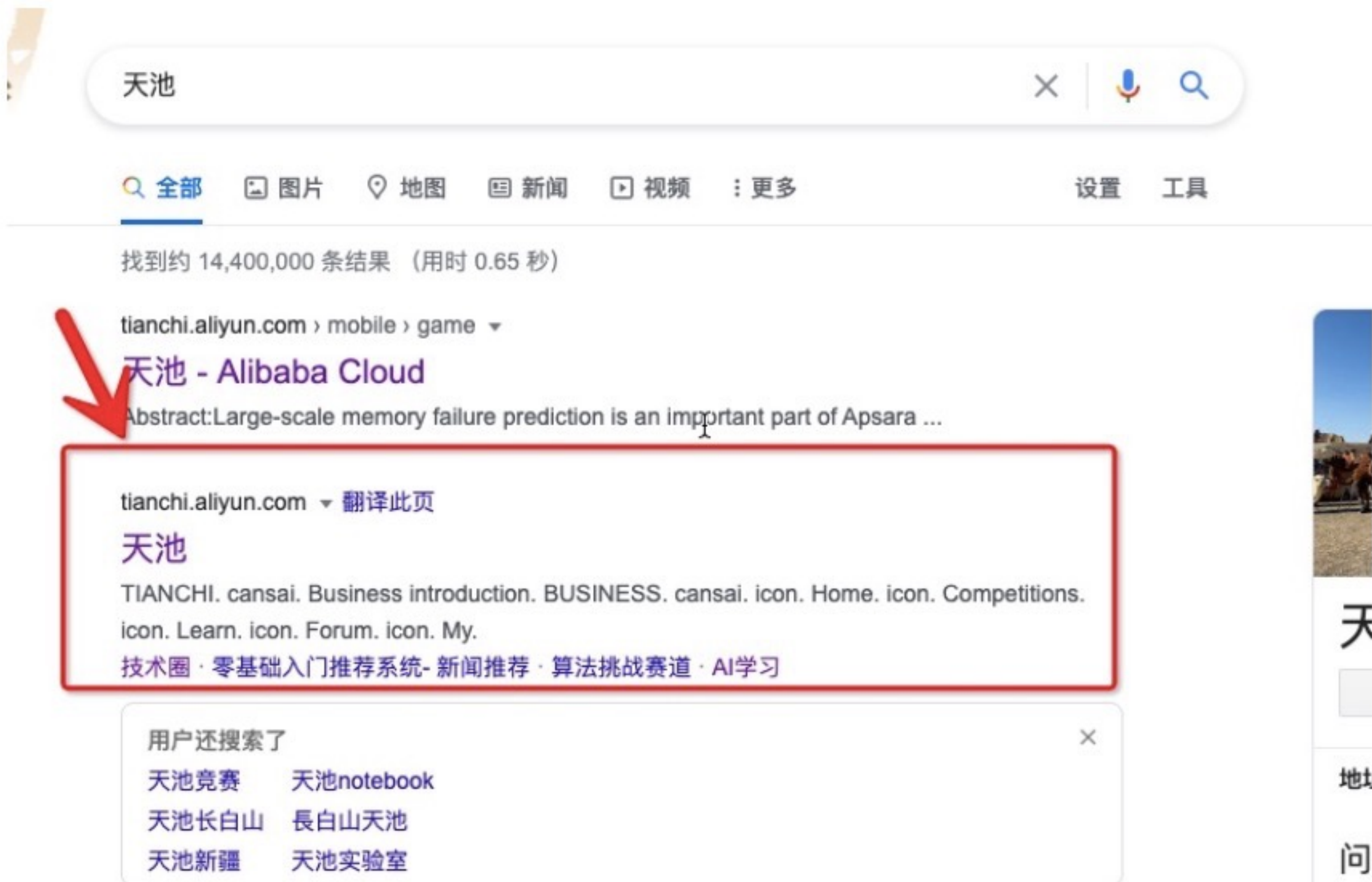
直播时间：6月22日20:00



扫码领取读书会配套学习资源



1) 首先需要进入天池官网，大家打开浏览器，搜索 天池，找到 tianchi.aliyun.com即可访问进入天池官



网；

2) 在天池官网，将鼠标移到 天池学习，即可出现下拉列表，点击 天池读书会，即可进入天池读书会的页面。



3) 在天池读书会页面，你可以对对应的读书会图书进行提问，优秀的提问还有机会获得赠书，还可以点击配套的训练营或者课程资源进入学习，还有点击实践代码获取读书会的项目实践的代码，跟着我一起进行项目实践和代码学习，同时还有很多其他的读书会，大家也可以观看举办过的读书会的回放，或者预约还没开始的读书会。



陈亮 七月在线AI Lab负责人

直播主题 《统计学习方法》

直播时间 2021年6月22日 20:00

学习资料 《人工智能的数学基础》课程

实践项目 无



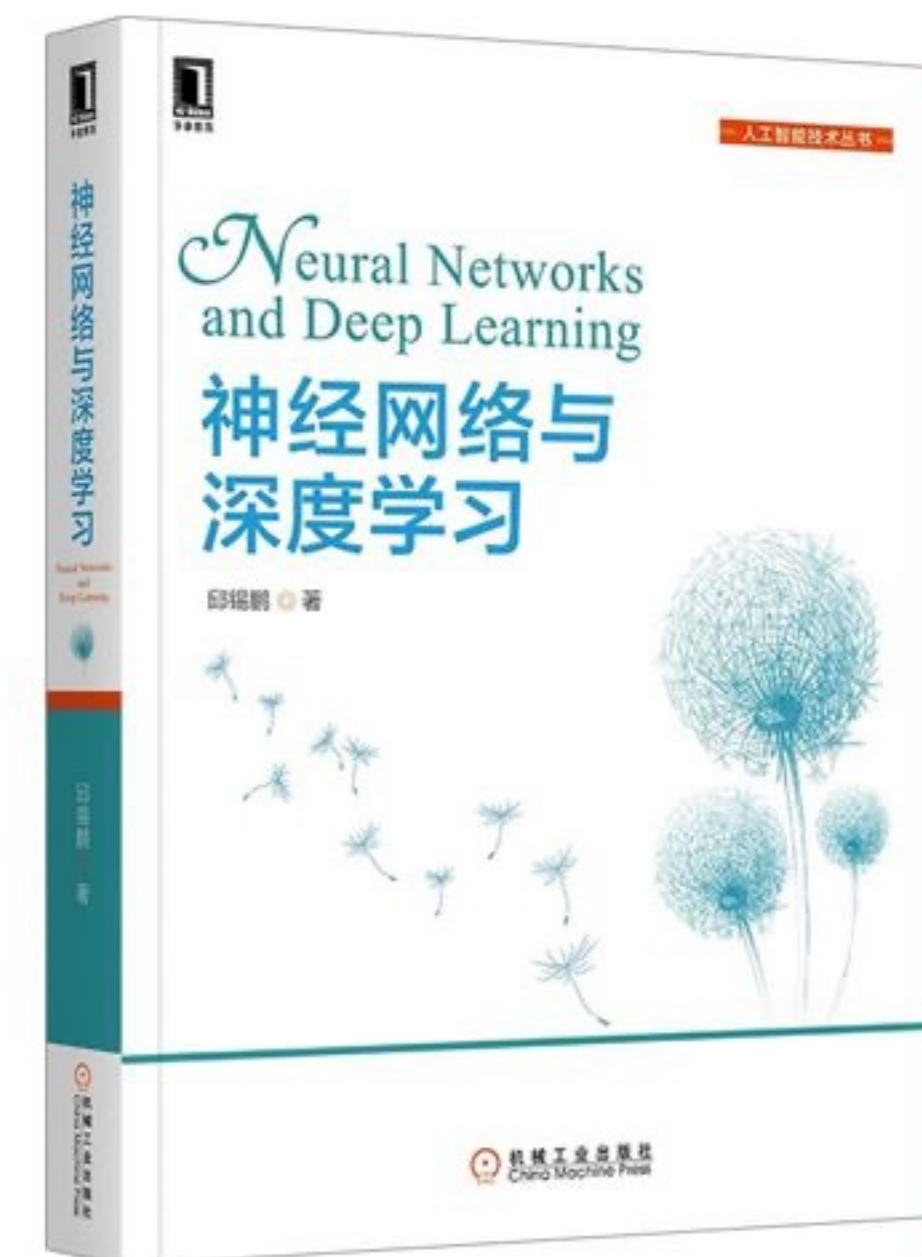
[🗨️ 提问](#) | [✍️ 学习课程](#) | [🛒 购买地址](#) | [📄 PPT下载](#) | [🕒 预约直播](#)

目录

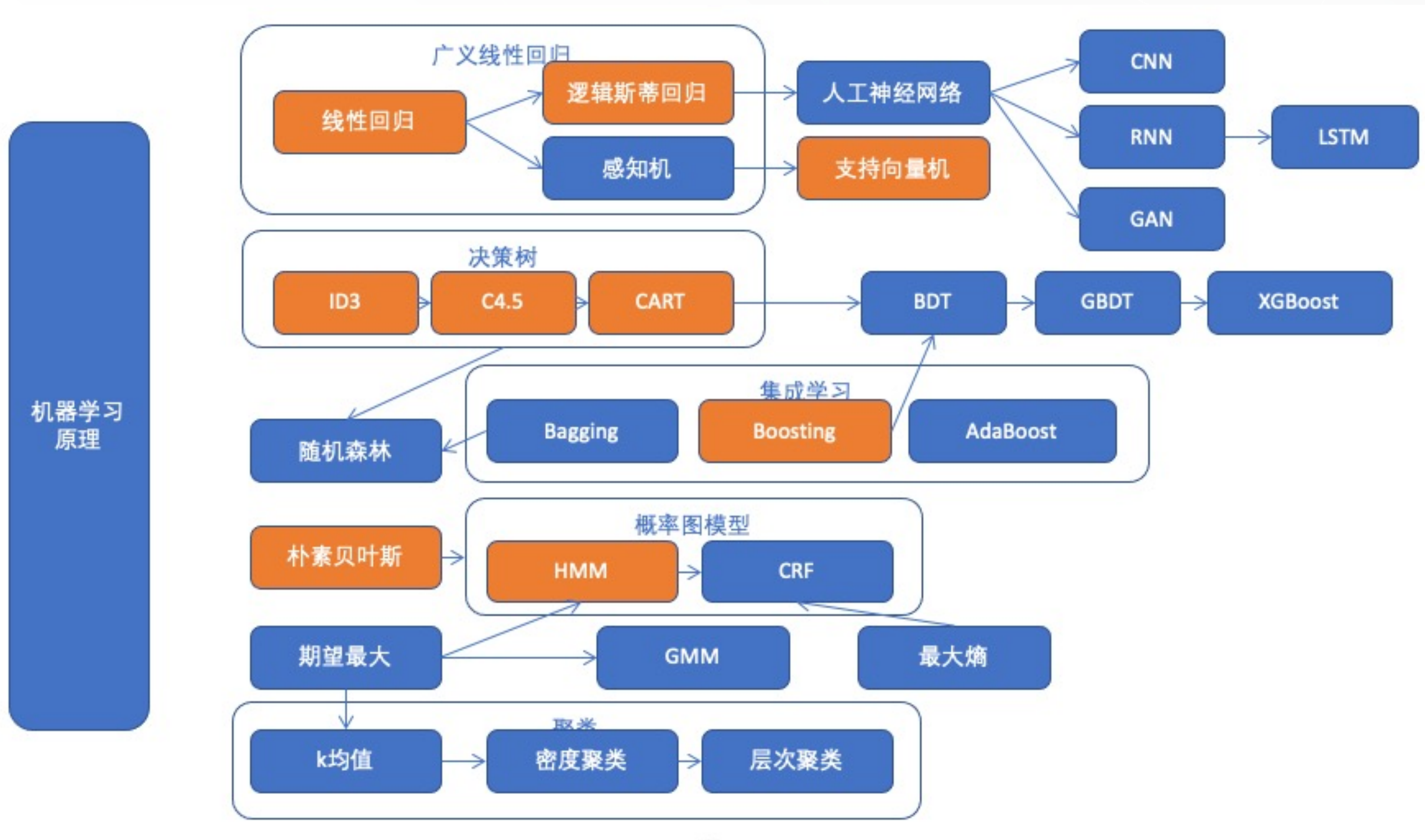
1. 图书简介
2. ML中的模型关联
3. 符号表及向量化
4. 决策函数与数形结合
5. 概率模型与基本规则



1. 图书简介



2. ML中的模型关联



3. 符号表及向量化

符号表

| | |
|---|------------------------|
| \mathbf{R} | 实数集 |
| \mathbf{R}^n | n 维实数向量空间, n 维欧氏空间 |
| \mathbf{H} | 希尔伯特空间 |
| \mathbf{X} | 输入空间 |
| \mathbf{Y} | 输出空间 |
| $x \in \mathbf{X}$ | 输入, 实例 |
| $y \in \mathbf{Y}$ | 输出, 标记 |
| X | 输入随机变量 |
| Y | 输出随机变量 |
| $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ | 训练数据集 |
| N | 样本容量 |
| (x_i, y_i) | 第 i 个训练数据点 |
| $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})^T$ | 输入向量, n 维实数向量 |
| $x_i^{(j)}$ | 输入向量 x 的第 j 分量 |
| $P(X), P(Y)$ | 概率分布 |
| $P(X, Y)$ | 联合概率分布 |
| \mathbf{F} | 假设空间 |
| $f \in \mathbf{F}$ | 模型, 特征函数 |
| θ, w | 模型参数 |
| $w = (w_1, w_2, \dots, w_n)^T$ | 权值向量 |
| b | 偏置 |
| $J(f)$ | 模型的复杂度 |
| R_{emp} | 经验风险或经验损失 |
| R_{exp} | 风险函数或期望损失 |
| L | 损失函数, 拉格朗日函数 |
| η | 学习率 |
| $\ \cdot\ _1$ | L_1 范数 |
| $\ \cdot\ _2, \ \cdot\ $ | L_2 范数 |
| $(x \cdot x')$ | 向量 x 与 x' 的内积 |

定义 2.1 (感知机) 假设输入空间 (特征空间) 是 $\mathcal{X} \subseteq \mathbf{R}^n$, 输出空间是 $\mathcal{Y} = \{+1, -1\}$. 输入 $x \in \mathcal{X}$ 表示实例的特征向量, 对应于输入空间 (特征空间) 的点; 输出 $y \in \mathcal{Y}$ 表示实例的类别. 由输入空间到输出空间的如下函数:

$$f(x) = \text{sign}(w \cdot x + b) \tag{2.1}$$

称为感知机. 其中, w 和 b 为感知机模型参数, $w \in \mathbf{R}^n$ 叫作权值 (weight) 或权值向量 (weight vector), $b \in \mathbf{R}$ 叫作偏置 (bias), $w \cdot x$ 表示 w 和 x 的内积. sign 是符号函数, 即

$$\text{sign}(x) = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases} \tag{2.2}$$

4. 决策函数与数形结合

假设空间用 \mathcal{F} 表示。假设空间可以定义为决策函数的集合：

$$\mathcal{F} = \{f|Y = f(X)\} \tag{1.5}$$

其中， X 和 Y 是定义在输入空间 \mathcal{X} 和输出空间 \mathcal{Y} 上的变量。这时 \mathcal{F} 通常是由一个参数向量决定的函数族：

$$\mathcal{F} = \{f|Y = f_{\theta}(X), \theta \in \mathbf{R}^n\} \tag{1.6}$$

参数向量 θ 取值于 n 维欧氏空间 \mathbf{R}^n ，称为参数空间 (parameter space)。

感知机有如下几何解释：线性方程

$$w \cdot x + b = 0 \tag{2.3}$$

对应于特征空间 \mathbf{R}^n 中的一个超平面 S ，其中 w 是超平面的法向量， b 是超平面的截距。这个超平面将特征空间划分为两个部分。位于两部分的点 (特征向量) 分别被分为正、负两类。因此，超平面 S 称为分离超平面 (separating hyperplane)，如图 2.1 所示。

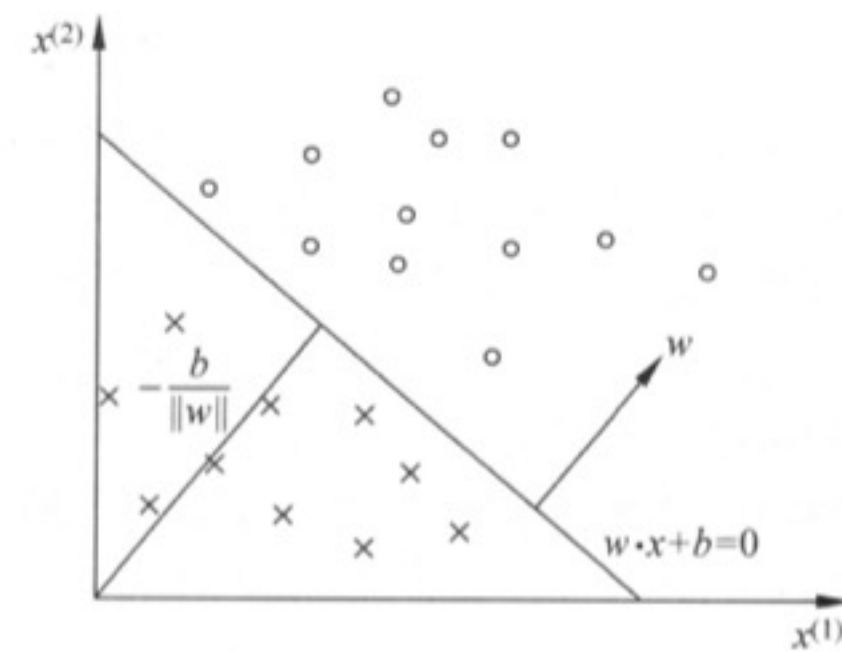


图 2.1 感知机模型

5. 概率模型与基本规则

假设空间也可以定义为条件概率的集合：

$$\mathcal{F} = \{P|P(Y|X)\} \tag{1.7}$$

其中， X 和 Y 是定义在输入空间 \mathcal{X} 和输出空间 \mathcal{Y} 上的随机变量。这时 \mathcal{F} 通常是由一个参数向量决定的条件概率分布族：

$$\mathcal{F} = \{P|P_\theta(Y|X), \theta \in \mathbf{R}^n\} \tag{1.8}$$

参数向量 θ 取值于 n 维欧氏空间 \mathbf{R}^n ，也称为参数空间。

监督学习分为学习和预测两个过程，由学习系统与预测系统完成。在学习过程中，学习系统利用给定的训练数据集，通过学习（或训练）得到一个模型，表示为条件概率分布 $\hat{P}(Y|X)$ 或决策函数 $Y = \hat{f}(X)$ 。条件概率分布 $\hat{P}(Y|X)$ 或决策函数 $Y = \hat{f}(X)$ 描述输入与输出随机变量之间的映射关系。在预测过程中，预测系统对于给定的测试样本集中的输入 x_{N+1} ，由模型 $y_{N+1} = \arg \max_y \hat{P}(y|x_{N+1})$ 或 $y_{N+1} = \hat{f}(x_{N+1})$ 给出相应的输出 y_{N+1} 。

加法规则： $P(x) = \sum_y P(x, y)$

乘法规则： $P(x, y) = P(x)P(y|x)$

其中 x 和 y 是随机变量

图 1.4 基本概率公式

前向概率

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, i_t = q_i | \lambda)$$

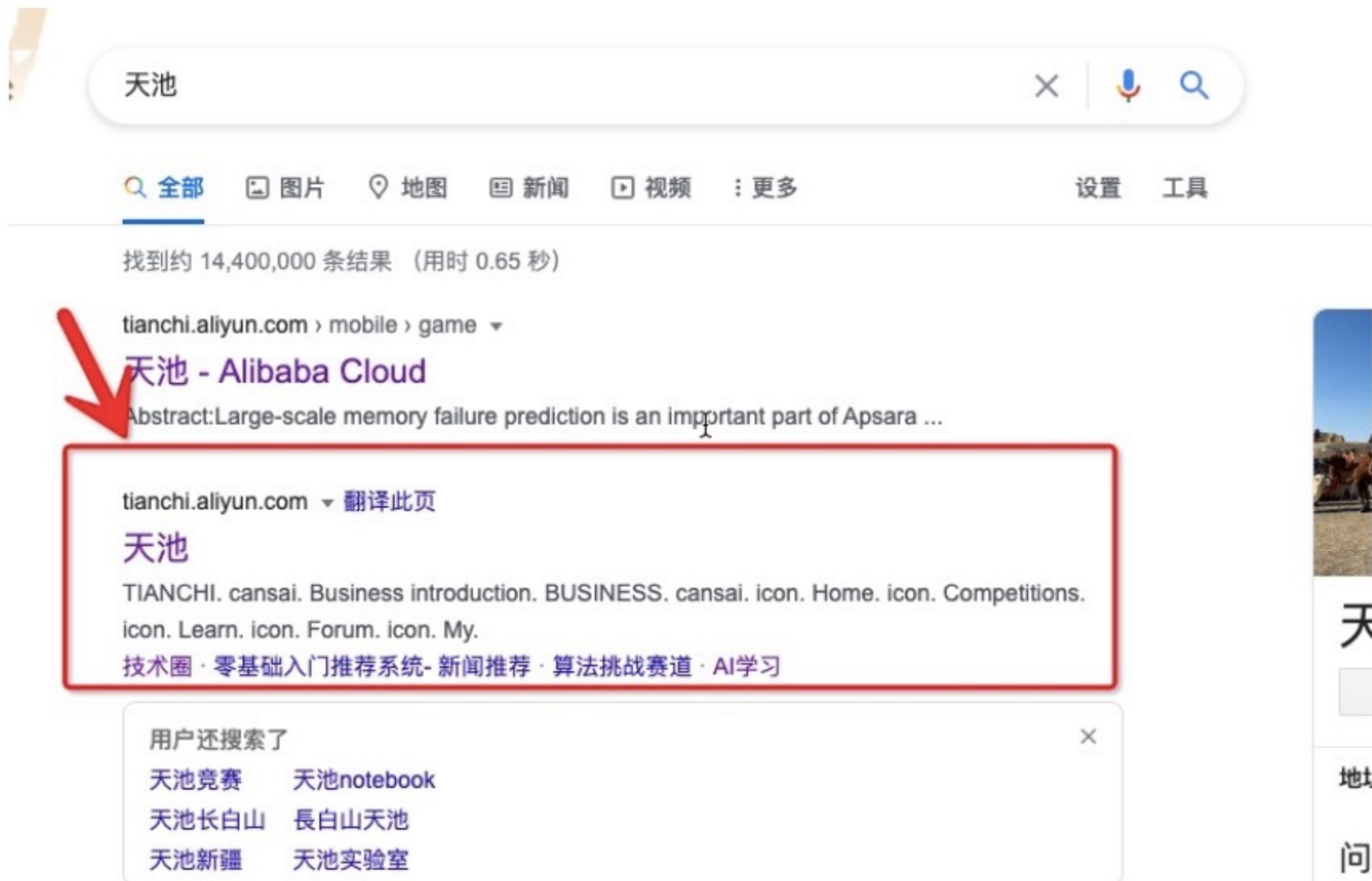
给定模型 λ ，时刻 t 部分观测序列为 o_1, o_2, \dots, o_t 且状态为 q_i 的概率。

前向概率递推计算

$$\begin{aligned} \alpha_t(i) &= P(o_1, o_2, \dots, o_t, i_t = q_i | \lambda) = P(i_t = q_i, o_t^i) \\ &= \sum_{j=1}^N P(i_{t-1} = q_j, i_t = q_i, o_t^i | \lambda) \\ &= \sum_{j=1}^N P(i_t = q_i, o_t^i | i_{t-1} = q_j, o_t^i) \cdot P(i_{t-1} = q_j, o_t^i | \lambda) \\ &= \sum_{j=1}^N P(i_t = q_i, o_t^i | i_{t-1} = q_j) \cdot \alpha_{t-1}(j) \\ &= \sum_{j=1}^N P(o_t^i | i_t = q_i, i_{t-1} = q_j) \cdot P(i_t = q_i | i_{t-1} = q_j) \cdot \alpha_{t-1}(j) \\ &= \sum_{j=1}^N b_i(o_t) \cdot a_{ji} \cdot \alpha_{t-1}(j) \end{aligned}$$

Q&A

1) 首先需要进入天池官网，大家打开浏览器，搜索 天池，找到 tianchi.aliyun.com即可访问进入天池官



网；

2) 在天池官网，将鼠标移到 天池学习，即可出现下拉列表，点击 天池读书会，即可进入天池读书会的页面。



3) 在天池读书会页面，你可以对对应的读书会图书进行提问，优秀的提问还有机会获得赠书，还可以点击配套的训练营或者课程资源进入学习，还有点击实践代码获取读书会的项目实践的代码，跟着我一起进行项目实践和代码学习，同时还有很多其他的读书会，大家也可以观看举办过的读书会的回放，或者预约还没开始的读书会。



陈亮 七月在线AI Lab负责人

直播主题 《统计学习方法》

直播时间 2021年6月22日 20:00

学习资料 《人工智能的数学基础》课程

实践项目 无



[🗨️ 提问](#) | [✍️ 学习课程](#) | [🛒 购买地址](#) | [📄 PPT下载](#) | [🕒 预约直播](#)

天池读书会

TIANCHI 天池



清华大学出版社
TSINGHUA UNIVERSITY PRESS

《统计学习方法》

从ML中的模型关联、符号表及向量化、决策函数与数形结合、概率模型与基本规则四个方面给大家分享本书。

直播嘉宾：陈亮 七月在线AI Lab负责人

直播时间：6月22日20:00



扫码领取读书会配套学习资源

