

机器学习中的数学符号与统计知识

基本符号

基本集合与逻辑

符号	LaTeX	读法	含义
\mathbb{R}	<code>\mathbb{R}</code>	实数集	所有实数的集合
\mathbb{R}^n	<code>\mathbb{R}^n</code>	n维实数空间	n维欧几里得空间
\mathbb{N}	<code>\mathbb{N}</code>	自然数集	所有自然数的集合
\mathbb{Z}	<code>\mathbb{Z}</code>	整数集	所有整数的集合
\in	<code>\in</code>	属于	元素属于集合
\subset	<code>\subset</code>	包含于	子集关系
\cup	<code>\cup</code>	并集	集合的并
\cap	<code>\cap</code>	交集	集合的交
\emptyset	<code>\emptyset</code>	空集	不包含任何元素的集合
\forall	<code>\forall</code>	对所有	全称量词
\exists	<code>\exists</code>	存在	存在量词
\wedge	<code>\wedge</code>	与	逻辑与
\vee	<code>\vee</code>	或	逻辑或
\neg	<code>\neg</code>	非	逻辑非
\Rightarrow	<code>\Rightarrow</code>	蕴含	逻辑蕴含
\iff	<code>\iff</code>	当且仅当	逻辑等价

向量与矩阵

符号	LaTeX	读法	含义
\mathbf{x}	<code>\mathbf{x}</code>	向量x	粗体小写字母表示向量
\mathbf{A}	<code>\mathbf{A}</code>	矩阵A	粗体大写字母表示矩阵
x_i	<code>x_i</code>	x下标i	向量x的第i个元素
A_{ij}	<code>A_{ij}</code>	A下标ij	矩阵A的第i行第j列元素
\mathbf{A}^T	<code>\mathbf{A}^T</code>	A转置	矩阵A的转置
\mathbf{A}^{-1}	<code>\mathbf{A}^{-1}</code>	A逆	矩阵A的逆
$\ \mathbf{A}\ $ 或 $\det(\mathbf{A})$	<code>\ \mathbf{A}\ </code> 或 <code>\det(\mathbf{A})</code>	A的行列式	矩阵A的行列式
$\text{tr}(\mathbf{A})$	<code>\text{tr}(\mathbf{A})</code>	A的迹	矩阵A对角线元素之和
\mathbf{I}	<code>\mathbf{I}</code>	单位矩阵	对角线为1, 其余为0的矩阵
$\mathbf{0}$	<code>\mathbf{0}</code>	零矩阵	所有元素都为0的矩阵
$\ \mathbf{x}\ $	<code>\ \mathbf{x}\ </code>	x的范数	向量x的长度或大小
$\ \mathbf{x}\ _2$	<code>\ \mathbf{x}\ _2</code>	x的L2范数	欧几里得范数, $\sqrt{\sum_i x_i^2}$
$\ \mathbf{x}\ _1$	<code>\ \mathbf{x}\ _1</code>	x的L1范数	$\sum_i x_i $
$\mathbf{x} \cdot \mathbf{y}$	<code>\mathbf{x} \cdot \mathbf{y}</code>	x点乘y	向量内积, $\sum_i x_i y_i$
$\mathbf{x} \otimes \mathbf{y}$	<code>\mathbf{x} \otimes \mathbf{y}</code>	x张量积y	向量的外积

概率论与统计

基本概率符号

符号	LaTeX	读法	含义
$P(A)$	<code>P(A)</code>	事件A的概率	事件A发生的概率
$P(A B)$	<code>P(A\ B)</code>	A条件于B的概率	在事件B发生的条件下，事件A发生的概率
$P(A, B)$	<code>P(A, B)</code>	A和B的联合概率	事件A和事件B同时发生的概率
$P(A \cup B)$	<code>P(A \cup B)</code>	A或B的概率	事件A或事件B发生的概率
$P(A \cap B)$	<code>P(A \cap B)</code>	A且B的概率	事件A和事件B同时发生的概率（同 $P(A, B)$ ）
$\mathbb{E}[X]$	<code>\mathbb{E}[X]</code>	X的期望	随机变量X的期望值
$\text{Var}(X)$	<code>\text{Var}(X)</code>	X的方差	随机变量X的方差， $\mathbb{E}[(X - \mathbb{E}[X])^2]$
σ_X	<code>\sigma_X</code>	X的标准差	随机变量X的标准差， $\sqrt{\text{Var}(X)}$
$\text{Cov}(X, Y)$	<code>\text{Cov}(X, Y)</code>	X和Y的协方差	$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$
$\rho_{X,Y}$	<code>\rho_{X,Y}</code>	X和Y的相关系数	$\frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$

常见分布

符号	LaTeX	读法	含义
$X \sim \mathcal{N}(\mu, \sigma^2)$	<code>X \sim \mathcal{N}(\mu, \sigma^2)</code>	X服从正态分布	X服从均值为 μ , 方差为 σ^2 的正态分布
$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	<code>f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}</code>	正态分布概率密度函数	均值为 μ , 方差为 σ^2 的正态分布密度函数
$X \sim \text{Bern}(p)$	<code>X \sim \text{Bern}(p)</code>	X服从伯努利分布	X服从参数为p的伯努利分布
$X \sim \text{Bin}(n, p)$	<code>X \sim \text{Bin}(n, p)</code>	X服从二项分布	X服从参数为n和p的二项分布
$X \sim \text{Pois}(\lambda)$	<code>X \sim \text{Pois}(\lambda)</code>	X服从泊松分布	X服从参数为 λ 的泊松分布
$X \sim \text{Unif}(a, b)$	<code>X \sim \text{Unif}(a, b)</code>	X服从均匀分布	X服从区间[a,b]上的均匀分布
$X \sim \text{Exp}(\lambda)$	<code>X \sim \text{Exp}(\lambda)</code>	X服从指数分布	X服从参数为 λ 的指数分布

机器学习中的重要概念

损失函数

名称	LaTeX	含义
均方误差 (MSE)	<code>L(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2</code>	$L(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
交叉熵损失	<code>L(\theta) = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]</code>	$L(\theta) = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$
L1正则化 (Lasso)	<code>\lambda \mathbf{w} _1 = \lambda \sum_{j=1}^p w_j </code>	$\lambda \mathbf{w} _1 = \lambda \sum_{j=1}^p w_j $
L2正则化 (Ridge)	<code>\lambda \mathbf{w} _2^2 = \lambda \sum_{j=1}^p w_j^2</code>	$\lambda \mathbf{w} _2^2 = \lambda \sum_{j=1}^p w_j^2$

梯度下降

梯度下降是机器学习中优化参数的基本算法，其更新规则如下：

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} L(\theta_t)$$

其中：

- θ_t 是第 t 步的参数
- η 是学习率
- $\nabla_{\theta} L(\theta_t)$ 是损失函数对参数 θ 的梯度

不同类型的梯度下降

类型	描述	公式
批量梯度下降	使用全部数据计算梯度	$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \frac{1}{n} \sum_{i=1}^n L(x_i, y_i, \theta_t)$
随机梯度下降	每次使用单个样本更新	$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} L(x_i, y_i, \theta_t)$
小批量梯度下降	使用批量样本计算梯度	$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \frac{1}{b} \sum_{i=1}^b L(x_i, y_i, \theta_t)$

神经网络相关

符号	LaTeX	含义
激活函数 Sigmoid	$\sigma(z) = \frac{1}{1 + e^{-z}}$	$\sigma(z) = \frac{1}{1 + e^{-z}}$
激活函数 ReLU	$f(z) = \max(0, z)$	$f(z) = \max(0, z)$
激活函数 tanh	$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$
Softmax函数	$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$	$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$
前向传播	$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}$	$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}$
激活值	$a^{(l)} = g(z^{(l)})$	$a^{(l)} = g(z^{(l)})$

评估指标

指标	LaTeX	含义
准确率	$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$	$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$
精确率	$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$	$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$

评估指标（续）

指标	LaTeX	含义
召回率	$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$	$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$
F1分数	$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
ROC曲线	-	接收者操作特征曲线，描绘不同阈值下真阳性率(TPR)与假阳性率(FPR)的关系
AUC	-	ROC曲线下的面积，值越接近1表示模型性能越好
均方根误差 (RMSE)	$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
平均绝对误差 (MAE)	$\text{MAE} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	$\text{MAE} = \frac{1}{n} \sum_{i=1}^n \ y_i - \hat{y}_i\ $
R^2分数	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

高级统计与机器学习概念

信息论

符号	LaTeX	读法	含义
$H(X)$	$H(X)$	X的熵	随机变量X的不确定性度量, $H(X) = - \sum_x P(x) \log P(x)$
$H(X Y)$	$H(X Y)$	X给定Y的条件熵	$H(X Y) = - \sum_{x,y} P(x,y) \log P(x y)$
$I(X;Y)$	$I(X;Y)$	X和Y的互信息	$I(X;Y) = H(X) - H(X Y) = H(Y) - H(Y X)$
$KL(P Q)$	$KL(P Q)$	P和Q的KL散度	$KL(P Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$
$JS(P Q)$	$JS(P Q)$	P和Q的JS散度	$JS(P Q) = \frac{1}{2} KL(P M) + \frac{1}{2} KL(Q M)$, 其中 $M = \frac{1}{2}(P + Q)$

贝叶斯方法

符号	LaTeX	读法	含义
$P(\theta D)$	<code>P(\theta D)</code>	后验概率	给定数据D的参数θ的概率
$P(D \theta)$	<code>P(D \theta)</code>	似然函数	给定参数θ观测到数据D的概率
$P(\theta)$	<code>P(\theta)</code>	先验概率	参数θ的先验信念
$P(D)$	<code>P(D)</code>	边际似然/证据	数据的边际概率, $P(D) = \int P(D \theta)P(\theta)d\theta$

贝叶斯定理:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

最优化理论

符号	LaTeX	读法	含义
$\nabla f(x)$	<code>\nabla f(x)</code>	f的梯度	函数f在点x处的梯度向量
$\nabla^2 f(x)$	<code>\nabla^2 f(x)</code>	f的海森矩阵	函数f在点x处的二阶导数矩阵
$\frac{\partial f}{\partial x_i}$	<code>\frac{\partial f}{\partial x_i}</code>	f对xi的偏导数	函数f对变量xi的偏导数
$\arg \min_x f(x)$	<code>\arg\min_x f(x)</code>	f的最小值点	使函数f取最小值的变量x的值
$\arg \max_x f(x)$	<code>\arg\max_x f(x)</code>	f的最大值点	使函数f取最大值的变量x的值

聚类与降维

符号/术语	LaTeX/描述	含义
K-means	K均值聚类	将数据集分为K个簇, 最小化样本到簇中心的距离平方和
$J = \sum_{j=1}^k \sum_{i:c_i=j} x_i - \mu_j ^2$	<code>J = \sum_{j=1}^k \sum_{i:c_i=j} x_i - \mu_j ^2</code>	K-means的目标函数, 最小化样本点到各自簇中心的距离平方和
PCA	主成分分析	通过线性投影减少数据维度的技术
SVD	奇异值分解	将矩阵分解为 $A = U\Sigma V^T$ 的矩阵分解方法
t-SNE	t-分布随机邻域嵌入	非线性降维技术, 保持数据点之间的相似性

高级机器学习模型

支持向量机(SVM)

支持向量机的优化目标：

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2$$
$$\text{subject to } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \forall i$$

带软边界的SVM（引入松弛变量）：

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$$
$$\text{subject to } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i$$

核函数：

核函数	LaTeX	含义
线性核	$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$	$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
多项式核	$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d$	$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d$
RBF核 (高斯核)	$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$	$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$
Sigmoid核	$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$	$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$

决策树与随机森林

术语	描述	公式
信息增益	特征分裂前后熵的减少	$IG(S, A) = H(S) - \sum_{v \in \text{Values}(A)} \frac{ S_v }{ S } H(S_v)$
基尼不纯度	衡量节点纯度的指标	$Gini(S) = 1 - \sum_{i=1}^c p_i^2$
随机森林	多个决策树的集成方法	通过投票或平均合并多个树的预测结果

深度学习高级概念

术语	描述	相关公式
反向传播	计算神经网络梯度的算法	$\frac{\partial L}{\partial w_{ij}^{(l)}} = a_j^{(l-1)} \delta_i^{(l)}$
Dropout	防止过拟合的正则化技术	训练时随机使部分神经元失活
BatchNorm	批量归一化	$\hat{x} = \frac{x - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$
Adam优化器	自适应学习率优化算法	结合动量和RMSProp的思想

强化学习

符号	LaTeX	含义
S	<code>S</code>	状态空间
A	<code>A</code>	动作空间
R	<code>R</code>	奖励函数, $R : S \times A \times S \rightarrow \mathbb{R}$
γ	<code>\gamma</code>	折扣因子, $0 \leq \gamma \leq 1$
π	<code>\pi</code>	策略, $\pi : S \rightarrow A$
$V^\pi(s)$	<code>V^{\pi}(s)</code>	状态价值函数, $V^\pi(s) = \mathbb{E}_\pi[\sum_{t=0}^\infty \gamma^t R_{t+1} S_t = s]$
$Q^\pi(s, a)$	<code>Q^{\pi}(s, a)</code>	动作价值函数, $Q^\pi(s, a) = \mathbb{E}_\pi[\sum_{t=0}^\infty \gamma^t R_{t+1} S_t = s, A_t = a]$
Bellman方程	-	$V^\pi(s) = \sum_a \pi(a s) \sum_{s'} P(s' s, a) [R(s, a, s') + \gamma V^\pi(s')]$

自然语言处理中的概念

术语	描述	相关公式
TF-IDF	词频-逆文档频率	$TF\text{-}IDF(t, d, D) = TF(t, d) \times IDF(t, D)$
Word2Vec	词嵌入模型	将单词映射到连续向量空间
RNN	循环神经网络	$h_t = f(h_{t-1}, x_t)$
LSTM	长短期记忆网络	解决RNN中的长期依赖问题的特殊RNN结构
Transformer	基于注意力机制的模型架构	$Attention(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$
BERT	双向Transformer编码表示	预训练语言模型

常用数学函数和符号

符号	LaTeX	读法	含义
$\sum_{i=1}^n x_i$	<code>\sum_{i=1}^n x_i</code>	x_i 从1到n的求和	对i从1到n的 x_i 求和
$\prod_{i=1}^n x_i$	<code>\prod_{i=1}^n x_i</code>	x_i 从1到n的连乘	对i从1到n的 x_i 求乘积
$\lim_{x \rightarrow a} f(x)$	<code>\lim_{x \rightarrow a} f(x)</code>	x 趋向于a时f(x)的极限	当 x 接近a时f(x)的极限值

符号	LaTeX	读法	含义
$\int_a^b f(x)dx$	<code>\int_a^b f(x) dx</code>	f(x)从a到b的定积分	函数f(x)在区间[a,b]上的定积分
$\frac{df}{dx}$	<code>\frac{df}{dx}</code>	f对x的导数	函数f关于变量x的导数
$\frac{\partial f}{\partial x}$	<code>\frac{\partial f}{\partial x}</code>	f对x的偏导数	多变量函数f关于变量x的偏导数
∇f	<code>\nabla f</code>	f的梯度	函数f的梯度向量，包含所有偏导数
∞	<code>\infty</code>	无穷大	表示无限大的数学符号
\approx	<code>\approx</code>	约等于	表示近似相等
\propto	<code>\propto</code>	正比于	表示两个量成正比关系
\sim	<code>\sim</code>	服从/分布于	表示随机变量服从某分布
\bar{x}	<code>\overline{x}</code>	x的均值	变量x的平均值
\hat{y}	<code>\hat{y}</code>	y帽	预测值或估计值y

时间序列分析

符号/术语	LaTeX/描述	含义
AR(p)	自回归模型	$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t$
MA(q)	移动平均模型	$X_t = \mu + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$
ARMA(p,q)	自回归移动平均模型	结合AR和MA模型的时间序列模型
ARIMA(p,d,q)	自回归积分移动平均模型	包含差分操作的ARMA模型
ACF	自相关函数	$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$
PACF	偏自相关函数	去除中间变量影响后的自相关

模型选择与交叉验证

术语	描述	相关公式
k折交叉验证	将数据分成k份，轮流使用其中一份作为测试集	-
留一法交叉验证 (LOOCV)	每次使用一个样本作为测试集的特殊交叉验证	-
网格搜索	通过遍历参数可能值的组合来寻找最优参数	-
AIC	赤池信息准则	$AIC = 2k - 2 \ln(\hat{L})$
BIC	贝叶斯信息准则	$BIC = k \ln(n) - 2 \ln(\hat{L})$
早停法	根据验证集性能提前结束训练的技术	-

异常检测与离群值处理

术语	描述	相关公式/方法
Z-score	标准分数	$z = \frac{x - \mu}{\sigma}$
IQR	四分位距	$IQR = Q_3 - Q_1$
离群值判定	基于IQR的常用判定规则	当 $x < Q_1 - 1.5 \times IQR$ 或 $x > Q_3 + 1.5 \times IQR$ 时，认为x是离群值
单类 SVM	用于异常检测的支持向量机变体	-
隔离森林	基于树的异常检测算法	根据样本被隔离的难易程度来判断异常
DBSCAN	基于密度的聚类方法，可用于异常检测	-

因果推断

术语	描述	相关概念
因果图	有向无环图(DAG)，表示变量间的因果关系	箭头表示因果方向
干预	$do(X = x)$	强制将X设为特定值x的操作
反事实	"如果...会怎样"	假设条件与实际情况不同时的结果
混淆因素	同时影响处理变量和结果变量的变量	-
工具变量	帮助识别因果效应的变量	满足特定独立性条件的变量
倾向得分	接受处理的条件概率	$P(T = 1 X = x)$

多臂老虎机与在线学习

术语	描述	相关算法/公式
探索-利用权衡	在已知策略(利用)和尝试新策略(探索)之间的平衡	-
ϵ -贪心	以概率 ϵ 随机选择动作，以概率 $1-\epsilon$ 选择当前最优动作	-
UCB算法	置信上界算法	$a_t = \arg \max_a \left(Q_t(a) + c \sqrt{\frac{\ln t}{N_t(a)}} \right)$
Thompson 采样	基于后验概率抽样来选择动作	-
再惩罚算法	通过计算最大遗憾来更新动作选择概率	-

元学习与迁移学习

术语	描述	常见方法
迁移学习	将一个任务学到的知识应用到另一个相关任务	微调、特征提取
领域适应	处理源域和目标域分布不同的问题	领域对抗网络
元学习	学习如何学习的方法	MAML、Reptile
少样本学习	从少量样本中学习的能力	原型网络、匹配网络
零样本学习	识别训练中未见过的类别	语义嵌入

数据预处理与特征工程

术语	描述	方法/公式
标准化	将特征缩放到均值为0，标准差为1	$z = \frac{x - \mu}{\sigma}$
归一化	将特征缩放到[0,1]区间	$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$
独热编码	将类别变量转换为二进制向量	-
特征选择	选择最相关的特征子集	过滤法、包装法、嵌入法
主成分分析(PCA)	降维技术，保留最大方差方向	通过特征值分解协方差矩阵
特征交叉	结合两个或多个特征创建新特征	如 $x_1 \times x_2$

概率图模型

术语	描述	相关概念
贝叶斯网络	有向无环图表示的概率模型	条件独立性
马尔可夫随机场	无向图表示的概率模型	团、势函数
隐马尔可夫模型(HMM)	观测值依赖于隐藏状态的概率模型	前向-后向算法
变分推断	近似计算复杂后验分布的方法	ELBO、KL散度
消息传递	在图模型中计算边缘概率的算法	信念传播、因子图

多目标优化

术语	描述	相关概念
帕累托最优	不能在不损害至少一个目标的情况下改进任何目标的解	帕累托前沿
加权和法	将多个目标函数加权组合成单一目标	$f(x) = \sum_{i=1}^m w_i f_i(x)$
约束法	优化一个目标，将其他目标作为约束	ϵ -约束法
进化多目标优化	使用进化算法求解多目标问题	NSGA-II、MOEA/D

常见的数学错误与术语

术语	描述	正确理解
相关性与因果性	相关并不意味着因果	观察到的相关性可能由第三个变量引起或纯属巧合
过拟合与欠拟合	模型复杂度与泛化能力的权衡	过拟合：模型过于复杂，捕捉噪声；欠拟合：模型过于简单，不能捕捉数据模式
偏差-方差权衡	模型偏差与方差之间的权衡关系	总误差 = 偏差 ² + 方差 + 不可约误差
P值的误解	P值不是假设为真的概率	P值是在原假设为真的条件下，观察到当前或更极端结果的概率

推荐系统相关概念

术语	描述	常用方法
协同过滤	基于用户-物品交互的推荐方法	基于用户的协同过滤、基于物品的协同过滤
内容过滤	基于物品特征的推荐方法	TF-IDF、主题模型
矩阵分解	将用户-物品交互矩阵分解为低维表示	SVD、NMF
冷启动问题	对新用户或新物品进行推荐的挑战	混合方法、基于内容的推荐

计算复杂度

符号	LaTeX	含义	例子
$O(f(n))$	<code>o(f(n))</code>	大O符号，表示算法复杂度的上界	$O(n^2)$ 表示算法最坏情况下复杂度不超过n的平方
$\Omega(f(n))$	<code>\Omega(f(n))</code>	大Omega符号，表示算法复杂度的下界	$\Omega(n)$ 表示算法最好情况下复杂度至少是n
$\Theta(f(n))$	<code>\Theta(f(n))</code>	大Theta符号，表示算法复杂度的紧确界	$\Theta(n)$ 表示算法复杂度与n成正比

- 基本数学符号（集合、逻辑、向量、矩阵）
- 概率论与统计学基础概念
- 常见概率分布及其表示
- 机器学习中的损失函数和评估指标
- 梯度下降及其变体
- 神经网络相关概念和公式
- 支持向量机、决策树、随机森林等经典算法
- 信息论与贝叶斯方法
- 最优化理论与聚类降维
- 深度学习高级概念
- 强化学习与自然语言处理
- 时间序列分析、异常检测和因果推断
- 多目标优化与元学习
- 计算复杂度表示