Tensor Wrold
人工智能讲座

人工智能的数学基础

AI工程师讲座

cangye@Hotmail.com

概率论 线性代数和矩阵 数学分析

目录

Part / 01

概率论

PROBABLITY TEHORY

概率论

PROBABLITY TEHORY

学习概率论的意义

算法复杂度分析

神经网络底层优化

EM算法

参数选取与优化

神经网络概率模型

信息论模型

概率论

PROBABLITY TEHORY

概率

$p(x)$

高斯分布

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{x^2}{2\sigma^2})$$

多变量高斯分布

$$p(\vec{x}_m) = (2\pi)^{-\frac{m}{2}} \left[\det\left(\underline{\underline{C}}\right)\right]^{-1/2} \exp(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \underline{\underline{C}}(\vec{x}-\vec{\mu})$$

条件概率

$p(x|\mathrm{A})$

“不上学的孩子容易成功”

先验概率

$p(x)$

“我从某渠道得知，印度人做生意80%都不诚实。所以我根据这个概率，选择不和印度人做生意。”

概率概念的应用：
高斯分布下的贝叶斯分类器

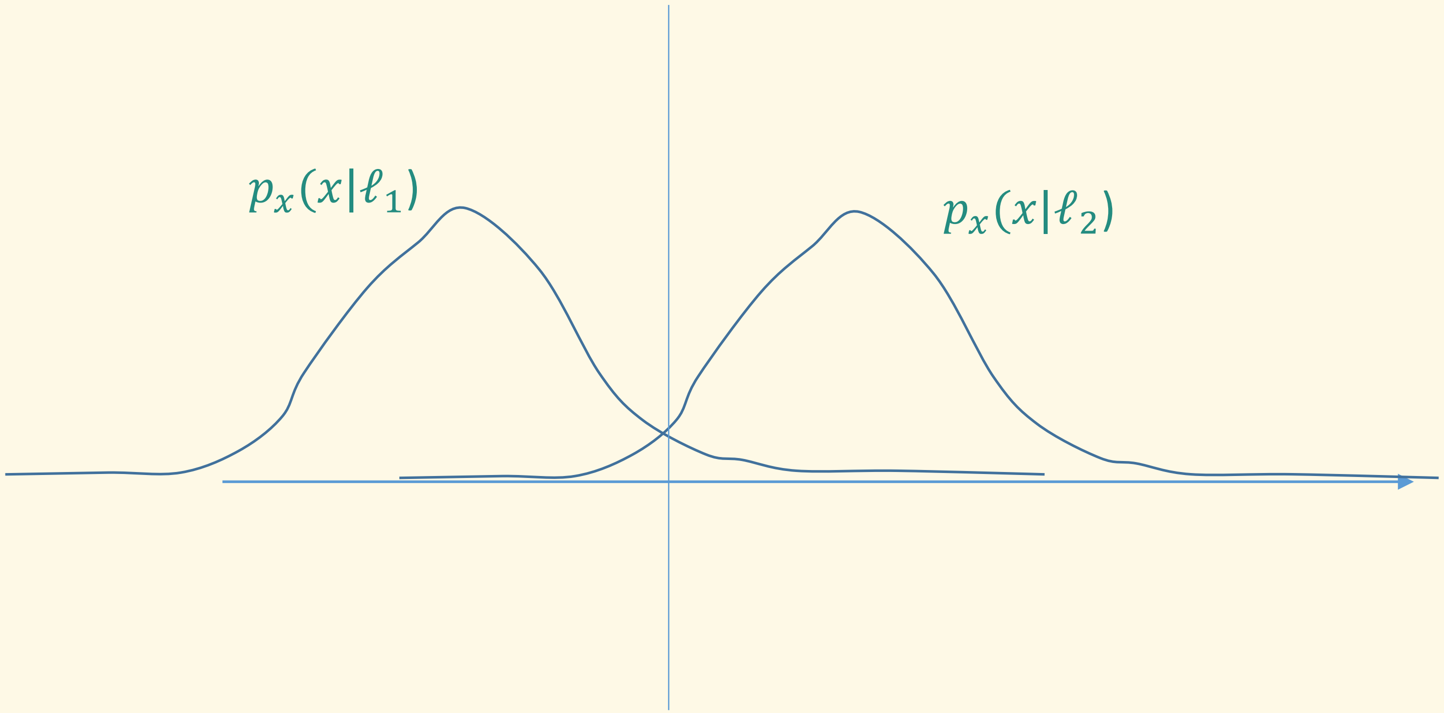
首先定义风险(几乎所有的机器学习分析都是从此开始的)：

$$\mathcal{R}$$

$$= c_{11}p_1 \int_{\mathcal{H}_1} p_x(x|\ell_1)\,dx + c_{22}\,p_2 \int_{\mathcal{H}_2} p_x(x|\ell_2)\,dx$$

$$c_{21}p_1 \int_{\mathcal{H}_2} p_x(x|\ell_1)\,dx + c_{12}\,p_2 \int_{\mathcal{H}_1} p_x(x|\ell_2)\,dx$$

高斯分布下的贝叶斯分类器等同于线性分类器



补充概念

变量均值

$$\mu = \mathbb{E}(x)$$

总体平均相关系数

$$\mathbb{E}(xx^T)$$

协方差矩阵(类比标量方差)

$$C = \mathbb{E}[(x-\mu)(x-\mu)^T]$$

补充概念

信息熵

$$H(x) = -\sum_{k=-K}^{K} p_k \log(p_k)$$

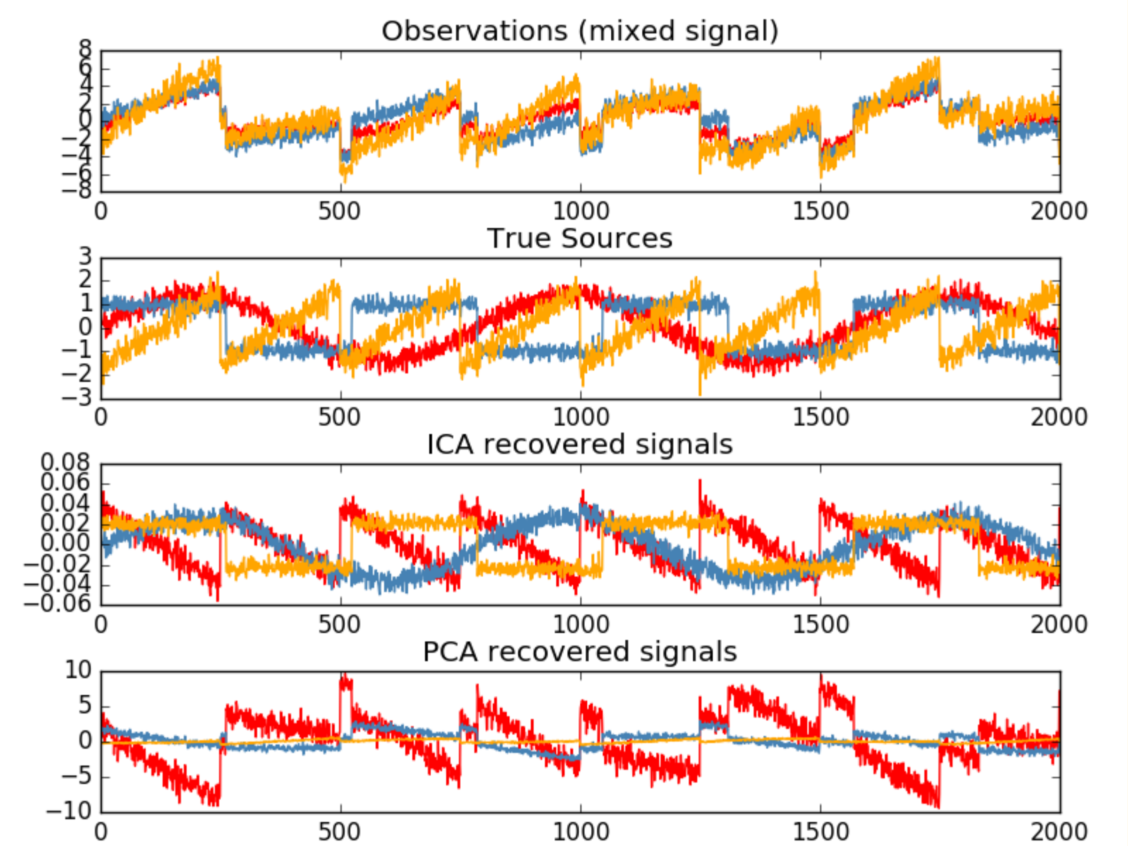
与热力学的熵定义相似

$$S = k ln\Omega$$

衡量系统的混乱程度

信息熵应用

ICA算法


Observations (mixed signal)
True Sources
ICA recovered signals
PCA recovered signals
0
500
1000
1500
2000

Part / 02

线性代数和矩阵

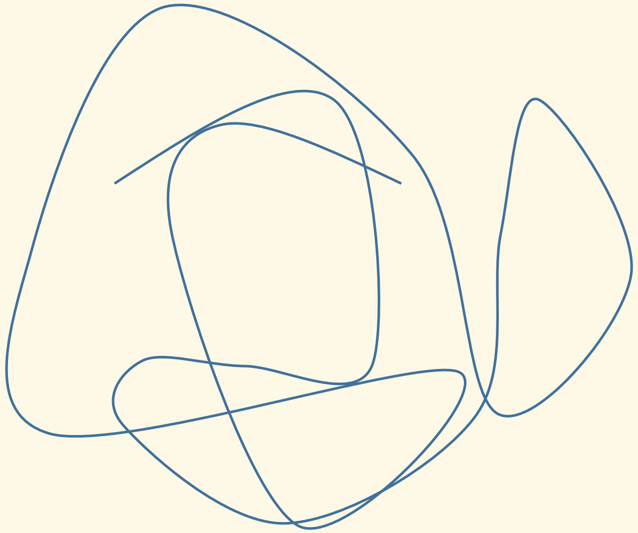
MATRIX

矩阵和张量

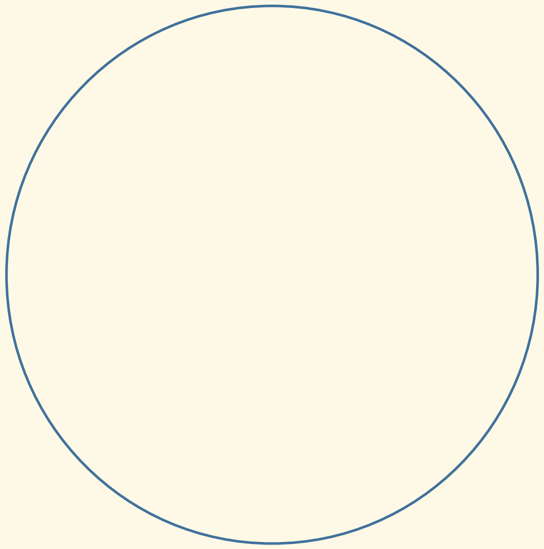
在机器学习中，张量的概念更接近于多维矩阵
而SVD分解等算法主要是针对于矩阵来说

在数学中，张量产生的目的在于表示不受坐标系变换的量，机器学习中的梯度等概念产生于数学中的张量。

张量以及衍生概念



左边我们在我们人工智能工作者的眼中称之为过拟合。
四维空间已经可以描述现象，十一维的弦理论依然有存在的意义



地心说描述的天体运动

日心说描述的天体运动

张量以及衍生概念

Hilbert Space

空间向量内积与二范数

$$x_i g_{ij} x_j = x_i x_i = ||\vec{x}||$$

N维度空间m维度超曲面

$$f_i(x_1, \ldots, x_n) = 0(i = 1 \ldots N - m)$$

为什么说这些？
回忆二维可分过程，一条线分两个类
三维空间中，一个面分两个类
…
N维空间中，N-1维曲面划分两个类

张量以及衍生概念

梯度与Hessian矩阵

在讨论梯度的时候我们本能的将其放在三维空间中想象，也就是说他是有“长度”的，所以是属于张量的概念。下图二维空间拉伸

张量以及衍生概念

梯度与Hessian矩阵

$f(\vec{x})$函数展开：

$$f(\vec{x}_0) + \nabla f(\vec{x})\Big|_{x=x_0} \cdot d\vec{x} + d\vec{x}\underline{\underline{H}}\Big|_{x=x_0} d\vec{x}$$

梯度：$\nabla f(\vec{x})$

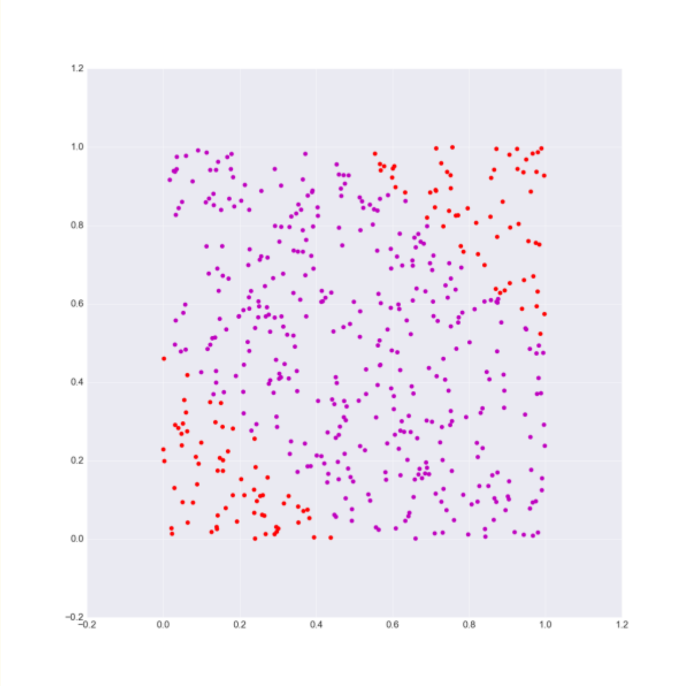
Hessian矩阵：$\underline{\underline{H}}$

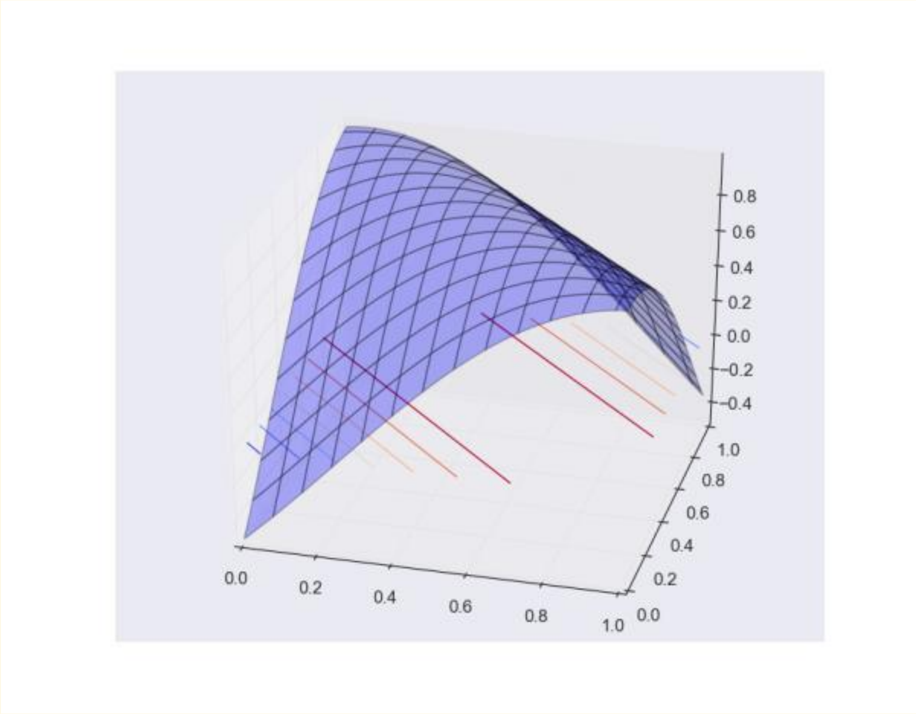
在将梯度写为$\nabla f(\vec{x}) = \frac{\partial f}{\partial x_i}$的时候

已经默认这是欧式空间中的梯度

张量以及衍生概念应用

回想前面说过的十一维
弦理论存在的意义
抑或问题





张量以及衍生概念应用

泛函

通常来讲我们会用一个数代表整个系统状态，而这个数字取最小(最大)值时系统进入稳定状态。则称系统为泛函。优化过程称之为凸优化。或者更直白一点，系统就是函数，泛函是函数的函数。

例子：
从家到公司走路需要消耗能量，如何找到一条能量最少的路。

矩阵的概念

矩阵的秩(rank)	线性独立的量的最小个数
矩阵的维(dimension)	与空间中概念类似
矩阵的迹(trace)	对角线元素之和

矩阵示例

放射变换

$$z = Ax + b$$

z, x为坐标，A为矩阵，这里矩阵的概念与张量无关

Part / 03

数学分析实例

ANALYSIS

本节例举一些简单的分析实例
用以加深数学原理的理解

贝叶斯分类器详述

风险函数：

$$\mathcal{R} = c_{22}p_2 + c_{21}p_1 + \int_{\mathcal{H}_1} [p_2(c_{12} - c_{22})p_x(x|\ell_2) - p_1(c_{21} - c_{11})p_x(x|\ell_1)]\, dx$$

最终化为比较Lambda大小

$$\Lambda = \frac{p_x(x|\ell_2)}{p_x(x|\ell_1)}$$

回忆：

$$p(\vec{x}_m) = (2\pi)^{-\frac{m}{2}}\left[\det\left(\underline{C}\right)\right]^{-1/2}\ \exp(-\frac{1}{2}(\vec{x}-\vec{\mu})^T\underline{C}(\vec{x}-\vec{\mu}))$$

最终化为

$$\log(\Lambda) = C^{-1}(\mu_1 - \mu_2)x + b$$

最小均方算法

风险函数：

$$\mathcal{R} = \left(d - f(x)\right)^2$$

风险函数：

$$\mathcal{R} = \left(d - f(x)\right)^2$$

R是关于x的凸函数，同时也是关于函数$f(x)$的泛函
这是泛函分析过程，构建凸函数是几乎所有有监督学习的起点。

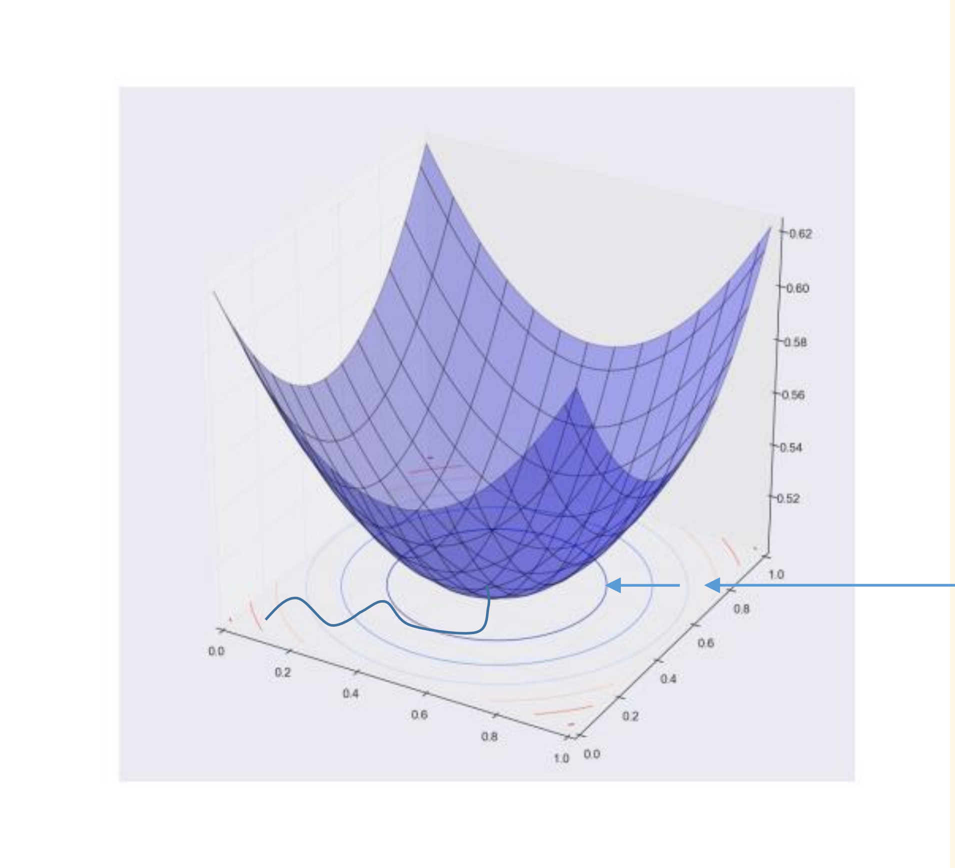
定义函数为

$$f(x) = Ax + b$$

所有的泛函数值求解过程都需要定义试函数，这通常是在无解析解的情况下所做的简化，现函数变为关于A的函数：

$$\mathcal{R}(A, b)$$

$\mathcal{R}(A,b)$最小值：



迭代过程：

$$x(n+1) = x(n) + \nabla_{A,b}\mathcal{R}$$

Tensor Wrold
人工智能讲座

THANKS

AI工程师讲座