

COMP0137: Machine Vision笔记

L.C.

注1: 每节课附上课件链接 (如果没有说明在讲上一节课的slide), 课程链接上面的是课堂上写的笔记, 链接下面的是Gemini根据课堂转录做的总结。

注2: 本文件旨在复习备考机器视觉, 无商用或其他目的, 并且内容均为L.C.本人主观角度出发编写, 应用人工智能辅助, 并且本人有三节课没有去, UCL Moodle也有多节课缺少录屏 (只能用去年的同一节课的录屏代替), 因此如有错误、偏见、短视等, 敬请谅解。

注3: 本文件最终解释权归L.C.所有, 请勿转载。

WEEK 1

2024.9.30 13.00-14.00

主要讲了整个课程的一些综述, 包括5月考试、2个作业、实验之类的, 然后讲了一些近邻的算法。教室在伦敦大学很偏, 很难评价。

课程参考书有两个, 主要用CVM, UDL在很后面有用到 (讲机器学习的那部分)

[Computer Vision Models, Learning, and Inference](#)

[Understanding Deep Learning.pdf](#)

平时有10个实验课作业

成绩: 线上开卷80分 (5月份, 选择题为主), coursework 2个各10分

[01 Introduction To MachineVision.pptx](#)

2024.10.1 9.00-10.00

网课 (老师开会去了), 讲的概率基础知识

[02 Introduction To Probability.pptx](#)

这节课是对**概率论基础知识**的回顾, 主要包括:

- 核心概念引入:** 通过实例 (如停止标志的颜色不完全一致) 引出不确定性的概念, 并介绍了**随机变量 (Random Variable)** (如 X, Y) 来表示这些不确定的量, 随机变量可以是离散的或连续的。
- 概率分布:** 描述随机变量取不同值的可能性大小, 包括离散分布 (如骰子点数) 和连续分布 (用概率密度表示, 如测试完成时间)。
- 关键概率类型:
 - 联合概率 $P(X, Y)$:** 多个随机变量同时取特定值的概率。
 - 边缘概率 $P(X)$:** 通过对联合概率 $P(X, Y)$ 中的其他变量 (如 Y) 进行求和或积分 (称为**边缘化 Marginalization**) 得到的单个变量的概率。
 - 条件概率 $P(X | Y)$:** 在给定某个变量 Y 取特定值的情况下, 另一个变量 X 取值的概率。
- 贝叶斯定理 (Bayes' Rule):** 基于条件概率和联合概率推导出的核心公式 $P(Y|X) \propto P(X|Y)P(Y)$, 解释了如何通过**先验概率 $P(Y)$** 和**似然函数 $P(X|Y)$** 计算**后验概率 $P(Y|X)$** 。

5. **独立性与期望**: 介绍了变量**独立 (Independence)** 的概念 (此时 $P(X|Y) = P(X)$) 以及**期望 (Expectation)** (概率加权平均值, 可用于计算均值、方差、协方差)。

2024.10.2 11.00-14.00

实验课, 主要就是要求配python环境和一些简单的python练习

2024.10.3 9.00-10.00

还是网课, 讲的进阶概率 (常见的几个分布还有先验)

[03_Probability_Distributions.pptx](#)

这节课主要讲解了**概率分布**作为参数化模型来表示数据。

核心内容包括:

1. **动机**: 对比了非参数方法 (如最近邻) 和参数化方法的优劣, 指出参数化模型需要选择合适的模型 (如直线或抛物线), 并承担模型选择带来的假设风险, 但可能更紧凑和易于泛化。
2. 常用分布介绍:
介绍了几种常用的概率分布及其适用场景, 包括:
 - **伯努利分布 (Bernoulli)**: 用于描述只有两个结果 (0或1) 的单次随机事件。
 - **贝塔分布 (Beta)**: 用于描述伯努利分布参数 λ 的不确定性, 其自身参数为超参数 α, β 。
 - **分类分布 (Categorical)**: 伯努利的推广, 用于描述有 K 个离散结果的单次随机事件。
 - **狄利克雷分布 (Dirichlet)**: 贝塔分布的推广, 用于描述分类分布参数 λ (一个向量) 的不确定性, 其自身参数为超参数 α (一个向量)。
 - **正态分布 (Normal/Gaussian)**: 用于描述连续变量。
 - 及其对应的参数分布 (如Normal Inverse Gamma, Normal Inverse Wishart)。
3. **共轭先验 (Conjugate Priors)**: 强调了上述分布经常成对出现 (如Beta-Bernoulli, Dirichlet-Categorical) 的原因是它们互为**共轭先验**。这意味着当先验分布 (如Beta) 与似然函数 (如Bernoulli) 结合时, 得到的后验分布仍然属于同一分布族 (如还是Beta分布, 只是参数更新了), 这极大地简化了贝叶斯推断中的计算。

WEEK 2

2024.10.7 13.00-14.00

线下课, 并且改教室了, seats直接没课了xs。。先是回顾了一下上周的网课内容, 然后这节课是拟合概率模型, 感觉就是机器学习和基础概率论的结合 (3种预测概率方法: 最大似然 - 最大后验 - 贝叶斯方法), 然后开始拟合正态分布, 提出可以对数化来简化运算

$$\begin{aligned} Pr(x_{1...I} | \mu, \sigma^2) &= \prod_{i=1}^I Pr(x_i | \mu, \sigma^2) \\ &= \prod_{i=1}^I \text{Norm}_{x_i}[\mu, \sigma^2] \\ &= \frac{1}{(2\pi\sigma^2)^{I/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^I \frac{(x_i - \mu)^2}{\sigma^2} \right] \end{aligned}$$

这节课主要将**概率论与机器学习中的模型拟合**联系起来。

核心内容包括：

1. **回顾与动机**：简要回顾了贝叶斯定理的术语（后验、似然、先验、证据），并通过雷达探测鸟群的例子引出问题：如何根据观测数据（如鸟的速度）建立模型。
2. 模型拟合的三种方法：
 - **最大似然估计 (Maximum Likelihood, ML)**: 寻找能使观测数据出现概率（似然函数）最大的模型参数 θ 。
 - **最大后验概率估计 (Maximum A Posteriori, MAP)**: 在最大似然的基础上，引入模型参数的先验分布 $P(\theta)$ ，寻找能使参数后验概率 $P(\theta | \text{Data})$ 最大的模型参数 θ 。
 - **贝叶斯方法 (Bayesian Approach)**: 不选择单一的最优参数，而是考虑所有可能的参数 θ ，并根据它们的后验概率进行加权。在预测时，通过积分（边缘化）消除对具体参数 θ 的依赖。
3. **实践考虑**：讨论了使用**对数似然 (log-likelihood)** 可以将乘积转化为求和，便于计算和优化，并提高数值稳定性。
4. **实例引入**：开始以拟合**正态分布**为例，具体演示如何应用这些方法（尤其是最大似然估计）。

2024.10.8 9.00-10.00

继续讲上一章 概率的东西讲了一会儿，然后开始MAP,逆伽玛，并且用贝叶斯计算后验概率

这节课继续深入探讨了**模型拟合**的三种方法（最大似然ML、最大后验MAP、贝叶斯），并以**拟合正态分布**到连续数据为例进行了详细推导：

1. **回顾与准备**：重申了三种方法的目标，并指出拟合正态分布（参数为均值 μ 和方差 σ^2 ）时，其共轭先验是正态-逆伽马分布。
2. **ML推导**：通过最大化对数似然函数，对 μ 和 σ^2 分别求偏导并令其为零，得到了ML估计的封闭解： μ_{ML} 是样本均值， σ_{ML}^2 是样本方差。
3. **MAP推导**：结合了正态分布的似然函数和正态-逆伽马先验，同样通过最大化对数后验概率（或其正比项），推导出了 μ_{MAP} 和 σ_{MAP}^2 的封闭解。结果显示MAP估计是样本统计量和先验信息的融合，且数据量越大，越接近ML估计。
4. **贝叶斯推导**：利用共轭性质指出，参数 (μ, σ^2) 的后验分布仍然是正态-逆伽马分布，只是其超参数 $(\alpha, \beta, \gamma, \delta)$ 根据数据进行了更新。这种方法保留了对参数的不确定性。
5. **贝叶斯预测**：预测新数据点时，贝叶斯方法通过对后验参数分布进行积分（边缘化），得到考虑了参数不确定性的预测分布，通常比MAP基于单点参数的预测更宽（更稳健）。

2024.10.9 11.00-14.00

实验课，做practical1abc，很简单，做完等了一会儿就溜了。

2024.10.10 9.00-10.00

从分类分布开始讲到狄利克雷分布，然后用似然（纯概率）来搞ML\MAP\Bayesian对比，然后开始下一章：

正态分布（单变量、多变量），然后讲了三种协方差，然后说到了变量的转换（旋转、投影）、边缘化高斯，变量变化/多重正态。

[05 Normal Distribution.pptx](#)

这节课主要包含两部分内容：

1. 完成分类分布 (Categorical Distribution) 的模型拟合实例：

- 详细推导了使用最大似然 (ML)、最大后验 (MAP) 和贝叶斯方法拟合分类分布（例如，动物园中不同动物出现的频率）的过程。
- ML 的结果是观测到的频率。MAP 和贝叶斯方法结合了观测数据和狄利克雷 (Dirichlet) 先验，其结果 (MAP点估计或贝叶斯后验均值) 是观测计数和先验伪计数的融合。
- 强调了贝叶斯方法能够更好地处理不确定性，例如即使某类别在本次观测中未出现，只要先验认为它可能存在，贝叶斯预测仍会给它分配一个非零概率。

2. 正态分布 (Normal/Gaussian Distribution) 的关键性质：

- 介绍了多元正态分布及其不同类型的协方差矩阵（球形、对角、全协方差）及其复杂度。
- 重点讲解了正态分布的几个重要数学特性：
 - **线性变换不变性：** 正态变量经过线性变换后仍是正态变量。
 - **边缘化不变性：** 多元正态分布的边缘分布（只看部分变量）仍是正态分布。
 - **条件化不变性：** 多元正态分布的条件分布（给定部分变量值，看剩余变量）仍是正态分布。
 - 这些性质及其对应的参数计算公式将在后续课程中频繁使用。

WEEK 3

2024.10.14 13.00-14.00

回顾了上周的内容，然后介绍了计算机视觉的都概念和解决方法，然后讲了三种CV的模型，接着详细分析了各情况下的回归任务。

[06 Learning And Inference.pptx](#)

这节课主要介绍了在计算机视觉 (CV) 中用于**建立观测 (X) 和世界状态 (W) 之间关系的概率模型**。

核心内容包括：

1. **动机与目标：** 以**背景减除**为例，说明了直接比较像素值会受光照变化、传感器噪声（像素抖动）等因素干扰，因此需要用概率模型（如高斯分布）来更鲁棒地描述背景。最终目标通常是求解 $P(W|X)$ ，即给定观测X，推断世界状态W的概率分布。
2. 三种模型类型：
 - **判别式模型 (Discriminative):** 直接对目标条件概率 $P(W|X)$ 进行建模。
 - **生成式模型 (Generative) - 联合概率:** 对联合概率 $P(X,W)$ 进行建模。
 - **生成式模型 (Generative) - 条件概率:** 对条件概率 $P(X|W)$ 进行建模（即给定世界状态生成观测）。
3. **建模流程：** 通常包括选择模型形式、选择学习算法（ML, MAP, Bayesian）拟合参数、以及进行推理。

4. **实例（回归问题判别式模型）**：开始讲解第一个实例——回归问题（ W 为连续变量，如姿态角度）。选择判别式方法，即直接建模 $P(W|X)$ 。具体选择用**高斯分布**来表示 $P(W|X)$ ，并假设高斯分布的**均值是输入 X 的线性函数**，而**方差是一个常数**。这个模型的参数（线性系数 ϕ_0, ϕ_1 和方差 σ^2 ）需要后续通过学习得到。

2024.10.15 9.00-10.00

继续讲了各情况下的分类任务，以及如何选取模型，为什么不用单一分布等等。

最后几分钟开了新一章的内容（多元高斯），并且提到在明天实验课继续讲一点。

[07 Modeling Complex Densities.pptx](#)

这节课主要围绕**分类任务**（即预测离散的世界状态 W ）展开，并引入了更复杂的概率模型：

1. **分类模型回顾**：对比了判别式模型（直接建模 $P(W|X)$ ，如使用Sigmoid函数）和生成式模型（建模 $P(X,W)$ 或 $P(X|W)$ ）。重点讨论了生成式模型 $P(X|W)$ 的情况：当观测 X 是连续的时，可以为每个离散类别 W 分别建立一个概率模型（如正态分布 $P(X|W=k)$ ），再结合先验 $P(W)$ 通过贝叶斯定理得到 $P(W|X)$ 。
2. **单一模型局限性**：通过实例指出，有时单一的简单分布（如一个高斯分布）不足以描述现实数据的复杂性（例如数据可能呈现多个峰值，即多模态分布）。
3. **混合模型引入**：为了解决单一模型的局限性，引入了**混合模型 (Mixture Models)** 的概念，特别是**高斯混合模型 (Mixture of Gaussians, MoG)**。MoG通过将多个高斯分布加权组合，可以拟合更复杂、具有多个峰值的概率密度。
4. **隐变量与学习**：解释了混合模型通常引入一个**隐变量 (latent variable)** 来表示数据点由哪个成分生成。学习混合模型的参数通常使用**期望最大化 (Expectation-Maximization, EM)** 算法，并预告了将在实验课中讲解和实践。

2024.10.16 11.00-14.00

实验课，前10多分钟接着昨天讲了高斯的内容，还有e-step，然后就做实验。。参考资料：

<https://moodle.ucl.ac.uk/mod/resource/view.php?id=6125156>

这节课（作为实验课的前导部分）主要讲解了使用**高斯混合模型 (Mixture of Gaussians, MoG)** 来**建模复杂概率密度**的核心概念：

1. **动机重述**：当单一高斯分布不足以描述数据的复杂性时（例如，数据呈现多个峰值，即**多模态 (multimodal)** 分布），需要更灵活的模型，如MoG。
2. **MoG模型结构**：MoG通过组合（加权求和）多个不同的高斯分布成分来形成一个更复杂的整体分布。它引入了一个**隐变量 (hidden variable)** h 来表示某个数据点是由哪个高斯成分生成的。
3. **学习算法 - EM**：由于隐变量的存在，无法直接用之前的封闭解方法（如最大似然的求导）来学习MoG的参数。需要使用**期望最大化 (Expectation-Maximization, EM)** 算法进行迭代学习。
4. EM算法步骤：
EM算法包含两个交替进行的步骤：
 - **E步 (Expectation)**: 计算每个数据点属于每个高斯成分的“**责任度 (responsibility)**”或后验概率 $P(h|X)$ 。
 - **M步 (Maximization)**: 使用E步计算出的责任度作为权重，重新估计（最大化）每个高斯成分的参数（均值、协方差）以及各成分的混合权重。

5. **实践要点：** EM算法保证每次迭代后数据的（对数）似然值不会下降。需要注意随机初始化的影响（可能需要多次运行取最优结果）以及可能出现的奇异性问题（某个高斯成分方差变得极小，集中于单个数据点）。

2024.10.17 9.00-10.00

继续在讲分类密度里面的混合高斯、T分布和因子分布的E-STEP和M-STEP 概率的东西。。还是第七章，讲了一半多吧。EM算法和应用好像就不讲了？

这节课继续讨论了**为复杂概率密度建模**的方法，重点在于介绍高斯混合模型 (MoG) 之外的其他利用隐变量的模型：

1. **回顾MoG与EM：** 再次提及高斯混合模型 (MoG) 是通过引入离散隐变量（选择哪个高斯成分）来处理多模态数据，并使用期望最大化 (EM) 算法进行迭代学习。解释了EM算法中通过优化似然函数的**下界 (lower bound)** 来保证迭代过程中似然值不下降。
2. 学生t分布 (Student's t-distribution):
 - **动机：** 解决高斯分布对**异常值 (outliers)** 敏感的问题。
 - **特性：** t分布具有比高斯分布更“重”的尾部，更能容忍远离均值的数据点。
 - **模型：** 可以看作是引入了一个连续的隐变量（服从Gamma分布）来控制高斯分布的方差（尺度），通过对该隐变量积分得到t分布。
3. 因子分析 (Factor Analysis):
 - **动机：** 解决**高维数据**（如图像像素）建模时，全协方差高斯模型参数过多的问题（维度灾难）。
 - **思想：** 假设高维观测数据 X 主要由少数几个潜在的**低维隐因子 (latent factors)** Z 线性生成，并叠加上噪声。
 - **模型：** 通过这种方式，可以用较少的参数（因子载荷矩阵 Φ 和噪声方差 Ψ ）来近似描述数据的协方差结构 ($\Phi\Phi^T + \Psi$)，实现降维。
4. **共性与组合：** 指出这些模型（MoG, t分布, 因子分析）都是通过引入不同类型的隐变量来增强模型表达能力，并且通常都使用EM算法进行参数学习。这些思想也可以组合使用（如t分布混合模型）。

WEEK 4

2024.10.21 13.00-14.00

回归: (让人头大的概率论)

线性回归: $Pr(w_i | \mathbf{x}_i, \boldsymbol{\theta}) = \text{Norm}_{w_i} [\phi_0 + \phi^T \mathbf{x}_i, \sigma^2]$

简化: $\mathbf{x}_i \leftarrow [1 \quad \mathbf{x}_i^T]^T$; $\phi \leftarrow [\phi_0 \quad \phi^T]^T$

计算整个数据集的概率（求积）：
$$Pr(\mathbf{w} | \mathbf{X}, \boldsymbol{\theta}) = \text{Norm}_{\mathbf{w}} [\mathbf{X}^T \phi, \sigma^2 \mathbf{I}]$$
$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_I] \quad \mathbf{w} = [w_1, w_2, \dots, w_I]^T$$

求似然：
$$\hat{\phi} = (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}\mathbf{w}$$
$$\hat{\sigma}^2 = \frac{(\mathbf{w} - \mathbf{X}^T \hat{\phi})^T (\mathbf{w} - \mathbf{X}^T \hat{\phi})}{I}$$

贝叶斯: $Pr(\phi | \mathbf{X}, \mathbf{w}) = \frac{Pr(\mathbf{w} | \mathbf{X}, \phi) Pr(\phi | \mathbf{X})}{Pr(\mathbf{w} | \mathbf{X})}$

$$Pr(\phi|\mathbf{X}, \mathbf{w}) = \text{Norm}_{\phi} \left[\frac{1}{\sigma^2} \mathbf{A}^{-1} \mathbf{X} \mathbf{w}, \mathbf{A}^{-1} \right]$$

$$\mathbf{A} = \frac{1}{\sigma^2} \mathbf{X} \mathbf{X}^T + \frac{1}{\sigma_p^2} \mathbf{I}$$

08 Regression.pptx

这节课主要讲解了**回归 (Regression)** 问题，特别是**线性回归**：

1. **问题设定**：回归任务是根据输入观测 X 预测一个或多个连续的输出值 W 。以从图像预测人体姿态（关节角度）为例。
2. **线性回归模型 (概率视角)**：将线性回归构建为一个判别式概率模型。假设给定输入 X ，输出 W 服从高斯分布，其均值是 X 的线性函数 ($\mu = \phi^T X$ ，通过技巧将截距项包含在 ϕ 和 X 中)，方差 σ^2 假设为常数。即 $P(W|X, \phi, \sigma^2) = N(W|\phi^T X, \sigma^2)$ 。
3. **最大似然估计 (ML)**：通过最大化训练数据的似然函数，可以推导出参数 ϕ 和 σ^2 的封闭解 (closed-form solution)，其中 ϕ 的解与经典的最小二乘法/标准方程结果一致。但ML方法可能在远离训练数据的区域过于自信（预测方差不变）。
4. **贝叶斯线性回归**：为了解决ML的局限性并融入先验知识，引入了贝叶斯方法。通常为参数 ϕ 设置一个高斯先验（如零均值先验，鼓励小参数值）。
5. **后验与预测**：利用贝叶斯定理，可以计算出参数 ϕ 的后验分布（由于共轭性，仍然是高斯分布）。在进行预测时，通过对参数 ϕ 的后验分布进行积分（边缘化），得到的预测分布不仅包含了数据本身的噪声 σ^2 ，还包含了模型参数 ϕ 的不确定性。这使得模型在远离训练数据的区域预测的不确定性（预测分布的方差）会相应增大。

2024.10.22 9.00-10.00

老师迟到了，以及只讲了大概40分钟？讲了非线性和高斯（核函数）还没讲完xs

这节课继续讨论**回归 (Regression)** 问题，重点从线性模型扩展到非线性模型：

1. **贝叶斯线性回归补充**：讨论了计算中的实用技巧，如当数据维度很高但样本数较少时，可以使用**矩阵求逆引理 (Matrix Inversion Lemma)** 来简化计算。同时提及了噪声方差 σ^2 通常可以通过最大似然法（在边缘化掉参数 ϕ 后）来估计。
2. **非线性回归引入**：为了拟合更复杂的数据模式（如曲线），引入了非线性回归。其核心思想是先对原始输入特征 X 进行**非线性变换**，得到一组新的特征 $Z=f(X)$ （这组变换函数称为**基函数 Basis Functions**，例如多项式基函数、高斯径向基函数RBF等），然后对变换后的特征 Z 应用**线性回归** 模型。
3. **核技巧 (Kernel Trick)**：指出在许多（特别是贝叶斯）回归算法的公式中，特征向量总是以内积（点积）的形式出现（如 $Z^T Z$ ）。核技巧允许我们定义一个**核函数 $*K(X_i, X_j)*$** ，它直接计算原始输入 X_i, X_j 对应的特征空间内积 $f(X_i)^T f(X_j)$ ，而无需显式地计算（可能非常高维甚至无限维的）特征变换 $Z=f(X)$ 。
4. **高斯过程回归 (Gaussian Process Regression)**：将核技巧应用于贝叶斯线性回归框架，就得到了高斯过程回归。这是一种强大的非线性贝叶斯方法，它直接在函数空间上定义先验，并且其预测结果能够自然地反映不确定性（在远离训练数据的区域不确定性增大）。不同的核函数（线性核、多项式核、RBF核等）对应了不同的函数先验（光滑度、周期性等）。

2024.10.23 11.00-14.00

practical3.都不难，线性、高斯、非线性回归，和解释原理。

2024.10.24 9.00-10.00

继续讲第8章。先完结了高斯部分，然后开始引入离散线性回归、对偶线性回归

这节课继续深入探讨**回归 (Regression)**，特别是处理**高维输入**和**防止过拟合**的方法：

1. **回顾与核函数**：简要回顾了使用核技巧 (Kernel Trick) 实现贝叶斯非线性回归（高斯过程回归），其中核函数的选择（如高斯核/RBF核）及其超参数（如RBF的宽度 λ ）会影响模型的拟合效果（可能欠拟合或过拟合）。
2. **稀疏性需求**：指出当输入特征 X 维度很高时，模型参数量可能过大，导致计算成本高且容易过拟合（模型可能学习到噪声或不相关的特征）。例如，用包含大量无关或歧视性问题的问卷来预测工作表现。
3. **稀疏线性回归 (Sparse Linear Regression)**：提出通过改变模型参数 ϕ 的先验分布来引入稀疏性。不再使用简单的高斯先验（鼓励所有参数都小），而是使用能鼓励**大部分参数为零**的先验（如学生t分布或拉普拉斯分布，后者对应L1正则化）。这相当于进行**自动特征选择**。实现上通常引入隐变量并使用EM算法。
4. **对偶表示 (Dual Representation)**：针对输入维度 d 远大于训练样本数 l 的情况，提出一种计算上更优的表示方法。将模型参数 ϕ ($d \times 1$ 维) 表示为训练样本 X ($d \times l$ 维) 的线性组合 $\phi = X\psi$ ，其中 ψ 是 $l \times 1$ 维的对偶权重。这样可以将原先涉及 $d \times d$ 矩阵的计算转换为涉及 $l \times l$ 矩阵的计算，在 $l \ll d$ 时大大降低计算量。
5. **相关向量机/回归 (Relevance Vector Machine/Regression, RVM/RVR)**：结合了稀疏先验和对偶表示的思想。它在对偶权重 ψ 上施加稀疏先验（如t分布），使得最终模型**仅依赖于少数几个“相关的”训练样本**（对应的 ψ_i 非零），这些样本被称为“相关向量”。RVM不仅能实现稀疏性，还能很好地结合核技巧处理非线性问题。

WEEK 5

2024.10.28 13.00-14.00

开始讲分类任务，首先是逻辑斯蒂、贝叶斯和非线性的回归

[09 Classification.pptx](#)

这节课主要讲解了**分类 (Classification)** 问题，特别是**逻辑回归 (Logistic Regression)** 模型及其优化方法：

1. **分类任务引入**：与上周的回归（预测连续值）不同，本周关注分类（预测离散类别标签 W ），首先从最简单的二分类问题入手。
2. **逻辑回归模型**：介绍了一种常用的判别式模型——逻辑回归。它通过一个线性函数 ($\phi^T X$) 计算“激活值”，然后将该激活值输入到 **Sigmoid 函数** 中，直接得到 $P(W=1 | X)$ 的概率。模型参数为 ϕ 。
3. **模型拟合挑战**：与线性回归不同，逻辑回归的最大似然估计**没有封闭解 (closed-form solution)**，不能通过简单求导置零直接解出最优参数 ϕ 。
4. **迭代优化方法**：因此，需要使用**迭代优化**算法来寻找最优参数。课程讨论了迭代优化的基本思想（从初始猜测出发，逐步改进以最小化代价函数或最大化似然函数），以及可能遇到的局部最优问题。
5. 两种优化算法：
 - **最速下降法 (Steepest Descent) / 梯度下降法 (Gradient Descent)**：只利用目标函数的一阶导数（**梯度**）来确定下降（或上升）最快的方向进行迭代更新。

- **牛顿法 (Newton's Method):** 同时利用目标函数的一阶导数 (**梯度**) 和二阶导数 (**海森矩阵 Hessian Matrix**)。它通过二次函数近似目标函数来确定更新步长和方向, 通常收敛速度更快, 但计算 (尤其是海森矩阵的求逆) 成本更高。

2024.10.29 9.00-10.00

继续讲分类, 讲了核、高斯进程、增量、拟合 (由于没有录屏, 这节课讲了啥存疑!)

这节课继续讨论**分类 (Classification)** 问题, 重点围绕逻辑回归的优化和贝叶斯处理方法:

1. **回顾逻辑回归优化:** 再次强调逻辑回归模型 (使用Sigmoid函数) 的最大似然估计没有封闭解, 必须采用迭代优化算法。回顾了基于一阶导数 (梯度下降/最速下降) 和二阶导数 (牛顿法) 的优化思路, 以及通过泰勒展开理解牛顿法 (用二次函数局部近似目标函数)。
2. **逻辑回归的凸性:** 指出逻辑回归的 (负) 对数似然函数是凸函数, 这意味着优化过程不会陷入局部最优, 保证可以找到全局最优解。
3. **贝叶斯逻辑回归引入:** 提出了逻辑回归的贝叶斯版本, 即为模型参数 ϕ 引入一个先验分布 (通常是高斯先验)。目的是为了正则化 (防止过拟合) 并能在预测中体现参数的不确定性。
4. **后验分布的挑战:** 指出贝叶斯逻辑回归的一个主要困难在于, 参数 ϕ 的后验概率分布 $P(\phi | \text{Data})$ 没有简单的解析形式 (即intractable), 因为它是由非共轭的似然函数 (伯努利/Sigmoid) 和先验 (高斯) 相乘得到的。
5. **拉普拉斯近似 (Laplace Approximation):** 为了解决后验分布 intractable 的问题, 介绍了拉普拉斯近似方法。该方法的核心思想是用一个**高斯分布**来近似真实的后验分布。这个近似高斯分布的**均值**取为真实后验分布的**峰值 (Mode)** (也就是MAP最大后验估计值), 其**协方差矩阵**由真实后验分布在峰值处的**负海森矩阵的逆** (即峰值处的曲率) 来决定。
6. **应用:** 得到的近似高斯后验分布可以用于后续的 (近似) 贝叶斯推理和预测, 使得在逻辑回归中也能考虑参数的不确定性。

2024.10.30 11.00-14.00

实验4 分类, 很简单。

顺便讲了coursework1

2024.10.31 9.00-10.00

继续分类, 讲了贝叶斯和增量问题。

这节课继续讨论**分类 (Classification)** 问题, 重点在**核方法**和**增量学习**:

1. **核逻辑回归 (Kernel Logistic Regression):** 回顾了如何使用核技巧 (如高斯核/RBF核) 将逻辑回归扩展到非线性情况。讨论了核函数超参数 (如RBF宽度 λ) 的重要性: λ 过小导致过拟合 (对训练数据拟合过好, 泛化能力差), λ 过大导致欠拟合 (模型过于简单, 无法捕捉数据复杂性)。强调了需要使用验证集来选择合适的超参数。
2. **贝叶斯核逻辑回归与稀疏性:** 简要提及了核逻辑回归的贝叶斯版本 (如高斯过程分类), 它可以更好地处理不确定性。进一步引入了**相关向量机分类 (Relevance Vector Classification, RVC)**, 它通过对偶空间中使用稀疏性先验 (如学生t分布), 能够得到一个**稀疏模型**, 即最终的决策仅依赖于少数几个“相关”的训练样本 (相关向量), 这有助于模型压缩和加速预测, 也利于后续的增量学习或在线学习。

3. **增量拟合/提升 (Incremental Fitting/Boosting)**: 作为构建复杂模型的另一种思路, 介绍了增量拟合方法。该方法并非一次性优化所有模型参数, 而是从一个简单模型开始, **逐步、贪婪地添加**新的简单模型 (称为“弱学习器”, 如带参数的非线性函数), 每次添加时主要关注之前模型未能正确分类的样本, 逐步提升整体性能。这种方法在计算上可能更高效, 尤其是在处理大规模数据或复杂模型时。

WEEK 6

2024.11.11 13.00-14.00

先讲了独立性有关的东西 (概率), 然后开始聚焦有向图的先验概率。举例介绍了几个 (有向图、无向图的对比)

[10 Graphical Models.pptx](#)

这节关于图模型的课程主要讲了:

1. **引入图模型**: 为了理解和简化大量变量之间的复杂关系 (比如团队协作中个体间的相互影响), 引入了图模型。
2. **核心概念 - 条件独立性**: 重点解释了条件独立性, 即在给定某些变量 (如 X_2) 的情况下, 其他变量 (如 X_1 和 X_3) 之间可能变得相互独立, 知道 X_3 对于预测 X_1 不再提供额外信息。
3. **图模型表示与优势**: 介绍了有向图模型 (贝叶斯网络), 其中变量关系用带箭头的边表示。这种结构允许将复杂的联合概率分布分解为一系列更简单的条件概率 (每个变量只依赖其父节点), 从而大大减少了表示和计算所需的参数数量与复杂度, 尤其是在变量数量庞大时。
4. **马尔可夫毯**: 解释了一个节点的马尔可夫毯 (包括其父节点、子节点、子节点的其他父节点), 给定马尔可夫毯内的节点, 该节点与其毯外的所有其他节点条件独立。

2024.11.12 9.00-10.00

还是继续在讲10章, 然后为了实验部分先跳到11章讲链的MAP 推理, 并类似讲了树。

[11 Models For Chains And Trees.pptx](#)

这节课接续图模型, 主要包括:

1. **图模型符号扩展**: 介绍了“盘子” (Plates) 表示法, 用于简洁地表示模型中重复出现的结构 (例如多个数据点或多个混合成分)。
2. **无向图模型**: 讲解了无向图模型 (马尔可夫网络), 它不依赖边的方向, 而是使用“势函数” (Potential Functions) 或等价的“代价/能量函数” (Cost/Energy Functions) 来定义变量子集 (称为“团”, Cliques) 之间的相互作用或兼容性, 并通过一个归一化常数 Z 来计算联合概率。
3. **链式模型的推断 (HMM与动态规划)**: 重点介绍了如何求解链状图模型 (如隐马尔可夫模型 HMM) 中的最可能状态序列问题 (Maximum A Posteriori, MAP)。课程展示了使用动态规划 (具体来说是维特比算法 Viterbi Algorithm) 的方法: 通过构建一个状态网格 (trellis), 从左到右迭代计算到达每个时间步每个可能状态的最小累积代价 (结合了观察代价“unary terms”和状态转移代价“pairwise terms”), 并记录路径, 最后从结尾回溯得到最优序列。
4. **效率**: 强调了动态规划相比于穷举搜索在计算效率上的巨大优势。

2024.11.13 11.00-14.00

今天的实验也比较短，好多都是选做（不做），大概就是一个链的视差值计算。

2024.11.14 9.00-10.00

今天把11、10章都结束了，把几种常见图的最大边界算法都详细介绍了。最后讲了一点12章的图算法。

这节课的内容可以总结为：

1. **回顾与深化：**回顾了上一节实验课的立体视觉（计算视差）例子，强调了动态规划在求解链状模型中寻找最大后验概率（MAP）解的作用，并说明了结合先验知识（如邻近像素视差相似）相比仅依赖似然（数据匹配）能得到更平滑、鲁棒的结果。
2. **树状图模型：**介绍了没有环路的树状图模型，这类模型也可以使用类似动态规划的消息传递算法（如信念传播，Belief Propagation，虽然具体算法细节可能未深入）进行高效精确推断。
3. 推断方法概览：
扩展了推断方法的工具箱，除了求解MAP解，还介绍了：
 - **最大边缘概率 (Max Marginals)：** 分别计算每个变量的边缘概率并取最大值，但这种方法可能导致全局次优解。
 - **采样方法：** 包括用于有向图的祖先采样 (Ancestral Sampling) 和用于无向图的马尔可夫链蒙特卡洛 (MCMC，如Gibbs采样)，这些方法通过从概率分布中抽取样本来近似推断，适用于连续变量或复杂模型，但不保证找到最优解。
4. **带环图的处理：**指出对于带环路的图模型（如网格），动态规划不直接适用，需要更复杂的精确算法（如提及了Graph Cuts可用于特定情况下的最优解）或近似推断方法。

WEEK 7

2024.11.18 13.00-14.00

终于不讲概率了。讲了小孔成像的原理（相机模型），还有一些数学上的投影和成像问题。

[14 The Pinhole Camera.pptx](#)

这节关于相机模型的课程可以总结为：

1. **引入针孔相机模型：**课程介绍了基本的针孔相机模型 (Pinhole Camera Model)，解释了三维世界中的点如何通过一个无限小的孔（光心 Optical Center）投影到成像平面 (Image Plane) 上，形成一个倒置反转的图像（通常用虚拟成像平面简化分析）。
2. 相机参数：
模型涉及两类参数：
 - **内部参数 (Intrinsics)：** 描述相机自身特性，包括焦距 (Focal Length f 或 ϕ ，可能x/y方向不同)、主点偏移 (Principal Point Offset Δx , Δy ，光心投影点与图像中心的偏差) 和像素倾斜/扭曲 (Skew γ)。
 - **外部参数 (Extrinsics)：** 描述相机在世界坐标系中的位姿，包括旋转 (Rotation Ω) 和平移 (Translation τ)。
3. **投影过程：**一个三维点从世界坐标系 w_world 投影到二维图像坐标 (x, y) 的过程分两步：首先通过外部参数（旋转和平移）将点变换到相机坐标系 w_camera ，然后通过内部参数和透视投影关系计算出最终的二维图像坐标。

4. **概率模型与镜头畸变**: 将投影过程建模为一个概率分布, 即预测的投影点是均值, 周围存在一定的不确定性。此外, 简要提及了真实镜头会引入径向畸变 (Radial Distortion), 虽然本次实验不考虑, 但在实际应用中需要校正。

2024.11.19 9.00-10.00

完了今天没去。。。据说接着前面一章讲了内参外参之类的。

这节课继续讨论相机模型, 主要内容总结如下:

1. 三个核心几何问题:

明确了基于针孔相机模型的三个基本计算机视觉任务:

- **求解外部参数 (Extrinsic Calibration / Exterior Orientation)**: 已知相机内参和3D点及其2D投影, 求解相机的旋转和平移 (位姿)。
- **求解内部参数 (Intrinsic Calibration)**: 已知相机外参和3D点及其2D投影, 求解相机的焦距、主点、倾斜等内参。
- **三维重建 (3D Reconstruction)**: 已知多个相机的内外参数和对应点的2D投影, 求解这些点的三维坐标。

2. **齐次坐标 (Homogeneous Coordinates)**: 为了简化计算, 引入了齐次坐标表示法。通过给2D/3D坐标增加一个维度, 可以将原本非线性的透视投影过程表示为线性的矩阵乘法形式 $\lambda * \tilde{x} = P * \tilde{w}$, 其中 P 是 3×4 的投影矩阵, \tilde{x} 和 \tilde{w} 分别是2D和3D点的齐次坐标。

3. **利用齐次坐标求解外部参数**: 详细演示了如何利用齐次坐标和已知的3D-2D点对应, 将求解外部参数 (旋转矩阵 Ω 和平移向量 t) 的问题, 转化为一个形如 $A * b = 0$ 的线性方程组。其中 b 是包含未知参数的向量, A 是根据已知点坐标构建的矩阵。

4. **SVD求解与约束**: 这个线性方程组可以通过奇异值分解 (SVD) 来求解 b 。得到的解还需要通过进一步处理 (对解中的旋转部分再做一次SVD并强制其满足旋转矩阵的性质) 来确保旋转矩阵的有效性, 并相应调整平移向量。

5. **预告**: 这种利用齐次坐标和线性代数求解的方法是后续解决其他问题 (包括下次实验的单应性计算 Homography) 的基础。

2024.11.20 11.00-14.00

实验课, 比较简单, 大概就是一个图像拼接, 然后后面还有相机投影和平面位姿估计点问题, 很多内容是重复的。很快就能做完, 最后有一个改进实验的部分。

还有一个糟糕的事情是这个课的课件消失了。。。本来应该是15, 现在moodle上只有15plus了。

这节课主要是为关于 **单应性 (Homography)** 的实验做准备, 内容可以总结为:

1. **引入单应性**: 单应性是一种特殊的变换, 描述了当所有三维点都位于同一个平面上时, 这些点如何投影到图像平面。这比处理任意三维点的情况要简单。
2. **数学表示**: 单应性变换可以用一个 3×3 的矩阵 Φ 来表示。在齐次坐标下, 从世界平面上的点 \tilde{w} (表示为 $[u, v, 1]^T$) 到图像上的点 \tilde{x} (表示为 $[x, y, 1]^T$) 的映射是线性的: $\lambda * \tilde{x} = \Phi * \tilde{w}$ 。这个 Φ 矩阵融合了相机的内参和外参 (针对平面场景)。

3. **求解方法**: 实验的核心任务是通过已知的点对应 (至少4对) 来计算单应性矩阵 Φ 。方法是利用齐次坐标表示的点 \tilde{x} 和变换后的点 $\Phi * \tilde{w}$ 应该共线 (方向相同) 的几何约束, 即它们的叉乘为零: $\tilde{x} \times (\Phi * \tilde{w}) = 0$ 。这个约束可以转化为关于 Φ 矩阵元素的线性方程组 $A * b = 0$ 。
4. **SVD求解**: 该线性方程组可以通过奇异值分解 (SVD) 求解, 从而得到单应性矩阵 Φ 的元素 (存储在向量 b 中)。
5. **应用与分解**: 计算出的单应性矩阵可以直接用于图像拼接 (如制作全景图)。如果需要, 并且已知相机内参, 还可以将 Φ 分解, 以估计相机相对于该平面的旋转和平移 (外部参数)。

2024.11.21 9.00-10.00

老师又迟到了麻了, 大致过完了transformation model的内容。开始从概率论变成线性代数了xs。

[15plus Models For Transformations.pptx](#)

这节课主要回顾和扩展了关于 **变换模型 (Transformation Models)** 的内容, 特别是针对平面场景和与单应性 (Homography) 的关系, 总结如下:

1. **变换模型层级与简化**: 介绍了处理平面场景的一系列变换模型, 包括欧几里得 (Euclidean)、相似 (Similarity)、仿射 (Affine), 它们都是单应性 (Homography) 的特例或简化形式。这些模型按复杂度 (自由度) 递增: 欧几里得 < 相似 < 仿射 < 单应性。相比处理任意三维场景的模型, 这些平面模型需要更少的点对应来求解。
2. **模型选择的重要性**: 强调了根据实际情况选择合适的模型复杂度。如果场景满足更简单的假设 (例如相机与平面平行且距离已知对应欧几里得变换), 使用更简单的模型可以用相同甚至更少数据获得更稳定、鲁棒的估计结果, 避免对噪声过拟合。
3. **统一的求解框架**: 再次确认了无论是求解这些简化的变换模型参数, 还是回顾求解相机内外参数或进行三维重建, 都可以利用齐次坐标将问题转化为线性方程组 (通常是 $A*b=0$ 形式), 然后使用奇异值分解 (SVD) 来求解未知参数。
4. 单应性的适用场景:

再次明确了单应性主要适用于两种情况:

- 相机观察一个 **平面场景** (此时相机可以任意移动)。
- 相机在 **原地进行纯旋转** (没有平移), 观察任意三维场景 (不一定是平面)。这是制作全景图 (图像拼接) 的技术基础, 因为纯旋转保证了不同视角下的图像之间存在单应性关系, 避免了因平移产生的视差问题。

WEEK 8

2024.11.25 13.00-14.00

时序模型, 没啥好说的, 最后好像开始讲卡尔曼滤波。

[19 Temporal Models.pptx](#)

这节关于时间序列模型 (Temporal Models) 的课程, 可以总结为:

1. **引入问题**: 课程以“猜豆子数量”和“估计兔子数量”为例, 引出了如何结合历史信息 (如昨天对兔子数量的高置信度估计) 和当前可能有噪声的观测 (如今天少数人对兔子数量的估计) 来推断随时间变化的状态。视频跟踪 (如跟踪车轮) 是另一个例子, 需要结合运动模型 (物理规律) 来处理遮挡、外观变化和相似物体干扰 (数据关联问题)。

2. **时间序列模型框架**：提出了一个通用框架来处理这类问题。系统有一个随时间 t 变化的隐藏状态 w_t （如位置、姿态、数量），以及与该状态相关的观测 x_t （如图像、传感器读数）。通常假设状态演化满足马尔可夫性（当前状态只依赖于上一状态），观测也只依赖于当前状态。
3. **递归贝叶斯估计（Recursive Bayesian Estimation）**：
解决这类问题的核心方法是递归地更新状态估计。这包含两个交替步骤：
 - **预测（Prediction / Time Update）**：使用 **时间模型**（描述状态如何随时间演化，如匀速运动模型、随机游走模型）和上一时刻的状态估计，来预测当前时刻的状态分布（先验概率）。预测通常会增加不确定性。
 - **更新（Update / Measurement Update）**：使用 **观测模型**（描述当前状态如何产生观测）和当前获得的新观测数据，来修正预测阶段得到的先验概率，得到当前时刻状态的后验概率分布。
4. **卡尔曼滤波器（Kalman Filter）介绍**：卡尔曼滤波器是实现递归贝叶斯估计的一种经典且高效的方法。它的关键 **假设** 是：时间模型和观测模型都是 **线性** 的，并且所有相关的噪声以及状态的先验/后验分布都是 **高斯分布**。在这种假设下，卡尔曼滤波器可以精确、快速地计算状态的均值和方差。课程图示了其基本循环：后验 \rightarrow 预测（高斯分布变宽） \rightarrow 结合观测似然（也是高斯） \rightarrow 更新得到新的后验（新的高斯分布）。

2024.11.26 9.00-10.00

还是在继续这一章，然后主要在讲卡尔曼滤波器和粒子滤波

这节课深入探讨了时间序列模型和滤波方法，可以总结为：

1. **卡尔曼滤波器回顾与深化**：详细回顾了卡尔曼滤波器（Kalman Filter）的预测（时间更新）和测量更新两个步骤的数学推导和含义。强调了其递归性质，即当前时刻的估计依赖于上一时刻的估计。通过实例展示了其滤波效果（平滑噪声、跟踪趋势但可能滞后）以及处理异步、不同精度传感器融合的能力。区分了实时滤波（Filtering）和利用未来信息的平滑（Smoothing）。
2. **卡尔曼滤波器的局限性**：指出标准卡尔曼滤波器的主要限制在于它 **假设** 状态转移模型和观测模型都是 **线性** 的，并且所有相关的噪声和状态分布都必须是 **高斯分布**。
3. **处理非线性问题**：
 - **扩展卡尔曼滤波器（Extended Kalman Filter, EKF）**：通过使用泰勒级数展开（计算雅可比矩阵）将非线性模型在当前估计点附近进行 **局部线性化**，从而扩展了卡尔曼滤波的应用范围。但当系统高度非线性时，这种线性化可能引入较大误差。
 - **无迹卡尔曼滤波器（Unscented Kalman Filter, UKF）**：通过一组精心选择的确定性采样点（Sigma Points）来传递均值和协方差信息，直接通过 **非线性函数** 进行变换，然后根据变换后的采样点重新估计高斯分布。通常比EKF在处理非线性时对协方差的估计更准确。
4. **处理非高斯/多峰分布：粒子滤波器（Particle Filter）**：
 - **动机**：当状态分布不再是简单的单峰高斯分布时（例如，目标可能在两个不同的位置），EKF/UKF 都会失效。
 - **核心思想**：使用大量带权重的 **随机样本（粒子）** 来近似表示任意形状的概率分布。
 - **基本流程（SIR - Sequential Importance Resampling）**：
通过迭代执行以下步骤进行状态估计：
 - **重采样（Resample）**：根据粒子权重重新抽取粒子，权重高的粒子被复制的概率更大。
 - **预测（Predict/Propagate）**：将重采样后的每个粒子通过（非线性）状态转移模型进行演化，并加入过程噪声。

- **更新 (Update/Weight):** 根据最新的观测数据，计算每个预测粒子与观测的 **似然度**，并将此似然度作为该粒子的新权重。
- **优点:** 非常灵活，可以处理 **非线性模型** 和 **非高斯噪声/多峰分布**。
- **缺点:** 计算量较大，在高维状态空间中可能效率低下（维度灾难）。

2024.11.27 11.00-14.00

实验课，07，粒子跟踪算法，A是因子化采样，估计后验概率分布；B是condensation凝聚算法，有点难做了两个小时效果很差放弃了。。（连续帧跟踪移动模板）

2024.11.28 9.00-10.00

印象很深的是上课放了一个什么矩阵之歌的视频，然后这节课还挺有意思但是跳着在讲

[16 Multiple Cameras.pptx](#)

这节课主要围绕 **基础矩阵 (Fundamental Matrix, F)** 和 **本质矩阵 (Essential Matrix, E)** 展开，它们是理解 and 处理两个摄像机视图之间几何关系的核心工具，尤其是在相机内外参数和三维点坐标都未知的情况下。总结如下：

1. **问题背景:** 当我们只有两个（或多个）不同视角的图像，但不知道相机的精确参数（内参、外参）或场景的三维结构时，如何从图像对应点中恢复这些信息，例如进行三维重建。
2. **对极几何 (Epipolar Geometry):** 这是两个视图之间内在的几何约束。一个视图中的点 x 在另一个视图中的对应点 x' 必须位于一条称为 **对极线 (epipolar line)** 的直线上。所有对极线会交于一点，称为 **对极点 (epipole)**。
3. **本质矩阵 (Essential Matrix, E):**
 - 描述的是 **已标定**（相机内参K已知）且 **归一化** 坐标下，两个视图对应点之间的关系。
 - 它包含了两个相机之间的相对旋转 R 和（单位化的）平移 t 信息，具体关系为 $E = [t]_{\times} * R$ （ $[t]_{\times}$ 是平移向量的反对称矩阵形式）。
 - 满足约束方程： $x_{\text{norm}}'^T * E * x_{\text{norm}} = 0$ 。
 - 具有5个自由度，秩为2。可以通过SVD分解来恢复 R 和 t （存在尺度和符号模糊）。
4. **基础矩阵 (Fundamental Matrix, F):**
 - 描述的是 **未标定**（相机内参K未知或不同）的相机像素坐标下，两个视图对应点之间的关系，是 **更通用** 的概念。
 - 它包含了本质矩阵E和两个相机的内参矩阵K、K' 的信息： $F = K'^{-T} * E * K^{-1}$ 。
 - 满足约束方程： $x_{\text{pixel}}'^T * F * x_{\text{pixel}} = 0$ 。
 - 具有7个自由度（9个元素 - 1个尺度因子 - 1个秩为2的约束），秩为2。
5. **八点算法 (Eight-Point Algorithm):**
 - 这是计算 **基础矩阵 F** 的常用线性方法。
 - 利用 $x'^T * F * x = 0$ 这个约束，每对匹配点 (x, x') 提供一个关于F矩阵9个未知元素的线性方程。
 - 至少需要 **8对** 匹配点，构建一个 $A * f = 0$ 的方程组（ f 是F矩阵元素的向量形式），通过 **SVD** 求解 f （取V矩阵的最后一列）。
 - 求得的F需要强制满足秩为2的约束（通过再次SVD并将最小奇异值置零）。

- 实际应用中常结合 **RANSAC** 来鲁棒地处理含有错误匹配 (outliers) 的数据。
6. **应用流程**: 计算出 F 后, 可以用来验证和滤除错误的特征匹配。如果已知相机内参 K , 可以计算 E , 进而分解得到相机间的相对姿态 R 和 t , 最后利用这些信息进行三维点的三角化 (Triangulation), 实现三维重建。

WEEK 9

2024.12.2 13.00-14.00

终于有点视觉了, 讲了几个基本的模型

[17 Models For Shape.pptx](#)

这节课主要介绍了计算机视觉中几种不同的二维形状表示和建模方法。

1. **核心内容**: 讨论了如何表示和分析图像中的物体形状, 强调了形状分析在计算机视觉实践中的重要性。
2. **主要模型**:
 - **蛇模型 (Snakes/Active Contours)**: 最灵活, 通过优化能量函数 (结合图像边缘信息和形状的平滑/连续性约束) 使轮廓点 (landmarks) 贴合目标边缘, 但缺乏全局形状知识。
 - **模板匹配 (Templates)**: 最不灵活, 使用一个固定的形状模板, 通过估计变换参数 (如平移、旋转、缩放、仿射变换等) 来匹配图像中的目标。参数少, 但适用性受限于变换模型。
 - **统计形状模型 (Statistical Shape Models, 如ASM, 概率PCA)**: 介于两者之间, 通过学习一组训练样本的形状变化 (如对齐后计算平均形状和主要变化模式/主成分), 建立形状的统计先验模型, 从而在匹配时既考虑图像特征, 又约束形状在合理范围内变化。
3. **关键概念**: 讨论了形状的定义 (去除尺度、旋转等因素后的几何信息)、界标点 (landmarks) 表示法、能量函数 (似然项+先验项)、距离变换、Procrustes对齐分析以及模型复杂性与灵活性的权衡。
4. **扩展**: 简要提及了这些概念可扩展到三维形状和纹理。

总的来说, 这节课从灵活但缺乏约束的蛇模型, 到刚性但参数少的模板匹配, 再到结合统计学习的更强大的统计形状模型, 系统地介绍了计算机视觉中处理形状问题的不同思路和技术。

2024.12.3 9.00-10.00

跨书了, 然后终于开始深度学习了xs。

[21 ShallowNeuralNets .pptx](#)

这节课的内容可以总结为以下几点:

1. **引入神经网络**: 课程开始从之前的概率模型过渡到神经网络, 特别是浅层神经网络, 并介绍了相关的学习资源 (新书、实验指导)。
2. **浅层网络结构**: 详细解释了一个简单的浅层神经网络如何工作: 输入信号先经过并行的线性变换 (权重乘以输入加偏置), 然后通过非线性激活函数 (如ReLU), 最后这些激活值再进行加权求和得到最终输出。
3. **功能与表示**: 说明了这种分层、组合的方式 (线性变换+非线性激活) 使得网络能够拟合复杂的非线性函数, 并将输入数据映射到一个 (可能更高维的) 特征空间, 这与之前讨论的特征工程、核方法、上下文模型等思想有关联。

4. **训练与应用铺垫**：简要提及了神经网络的训练方法——反向传播（Backpropagation），用于根据输出误差调整网络参数。同时介绍了一种无监督学习的应用——自编码器（Autoencoder），它通过让网络学习重构输入，在中间的“瓶颈”层得到输入的压缩表示。

总的来说，这节课通过图示和公式讲解了浅层神经网络的基本构成、工作原理及其表示能力，并为后续的实验课和更深入的学习（如反向传播、自编码器）打下了基础。

2024.12.4 11.00-14.00

实验08，神经网络，有 [MV Practical NN Cribsheet.pdf](#) 参考教材

第一部分 (Practical10a) 要求实现多层感知机 (MLP) 的关键组件，包括 ReLU 激活函数、仿射层和交叉熵 softmax 层，以及反向传播的梯度下降步骤。

第二部分 (Practical10b) 是使用预训练的 7 层自编码器，探索潜在空间，需要生成随机潜在编码，通过 PCA 找到主成分空间，并观察如何在潜在空间中平滑地在不同数字之间插值。

2024.12.5 9.00-10.00

有点招笑，今天才讲MLP，说白了这个课就是概率论最后加了一点点机器学习难绷

[22 MLPs Training_.pptx](#)

这节课的内容可以总结为以下几点：

1. **深入MLP结构与功能**：回顾了多层感知机 (MLP) 的基本单元和术语，并详细图解了对于1维和2维输入，带有ReLU激活函数的MLP如何通过产生“折点”或“折叠面”来将输入空间划分为多个线性区域，从而拟合复杂函数。
2. **网络深度的威力**：重点阐述了增加网络层数（深度）的意义。深层网络通过逐层对前一层的输出进行再划分和变换（函数的复合），能够以比单纯增加单层神经元数量（宽度）更高效的方式，构建出表达能力更强、更复杂的模型。
3. **训练挑战：非凸优化**：讨论了神经网络训练的核心——基于梯度的优化。通过例子说明损失函数通常是非凸的，存在多个局部最小值和鞍点，导致梯度下降法的结果很大程度上取决于参数的随机初始化，不一定能找到全局最优解。
4. **实用优化算法：SGD与动量**：介绍了随机梯度下降 (SGD)，即每次只用一小批 (mini-batch) 数据计算梯度来更新参数，以提高训练效率并可能跳出局部最优。进一步引入了动量 (Momentum) 等改进方法，通过累积历史梯度信息来加速收敛并稳定训练过程。

总的来说，这节课深入探讨了MLP的函数逼近能力、深层网络相比浅层网络的优势（表达效率）、训练过程中面临的非凸优化挑战以及常用的随机梯度下降及其改进算法（如动量法）。

WEEK 10

2024.12.9 13.00-14.00

不好意思没去，本人复习MLVC去了

[23 TrainingVsGeneralizing_.pptx](#)

这节课的内容可以总结为以下几点：

1. **评估与过拟合**：课程核心是如何衡量神经网络的性能，并着重讨论了训练过程中常见的“过拟合”问题，即模型在训练集上表现很好，但在未见过的测试集上表现差，甚至性能随着训练深入反而下降。
2. **偏差-方差权衡**：解释了过拟合背后的理论——偏差 (Bias) 和方差 (Variance) 的权衡。模型复杂度过低会导致高偏差 (欠拟合，模型无法捕捉数据规律)，复杂度过高则易导致高方差 (过拟合，模型对训练数据噪声过于敏感)。
3. **验证集的作用**：强调了在训练中必须使用独立的验证集 (Validation Set) 来监控模型的泛化能力。通过观察验证集上的性能变化，可以判断是否出现过拟合，并据此采取策略 (如早停法 Early Stopping, 选择最佳模型)。测试集 (Test Set) 则应保留到最后，用于对最终选定模型进行一次评估。
4. **应对策略简介**：指出了解决过拟合、平衡偏差与方差的几种途径，包括获取更多数据、调整模型复杂度 (容量 Capacity)，以及采用正则化技术。特别提到了数据增强 (Data Augmentation) 作为一种常用的正则化方法，通过对训练样本进行随机变换 (如添加噪声、旋转、平移等) 来扩充数据集，提高模型的鲁棒性。

总的来说，这节课从评估模型性能出发，深入讲解了过拟合现象、其背后的偏差-方差理论，并介绍了使用验证集监控训练以及通过数据增强等手段来改善模型泛化能力的关键实践方法。

2024.12.10 9.00-10.00

不好意思没去，本人复习MLVC去了 (好像讲了CNN)

[24 CNNs.pptx](#)

这节课的内容可以总结为以下几点：

1. **引入卷积神经网络 (CNN)**：针对全连接网络 (MLP) 处理图像时参数量巨大、忽略像素空间相关性 (如邻近像素值相似) 以及缺乏平移不变性/等变性 (图像移动后，识别结果应不变或相应移动) 等问题，引入了卷积神经网络。
2. **卷积层核心机制**：CNN的核心是卷积层。它使用小的滤波器 (卷积核 Kernel)，这些滤波器包含一组权重。滤波器在输入数据 (如图像) 上滑动，每次覆盖一小块区域，计算滤波器权重与该区域输入的加权和 (点积)，得到一个输出值。关键在于，同一个滤波器 (同一组权重) 在整个输入上是**共享的**，这大大减少了需要学习的参数数量。
3. **卷积操作细节**：讲解了卷积操作中的一些关键概念，如处理边界的填充 (Padding, 如补零或只计算有效区域)、控制输出大小和计算量的步长 (Stride)、处理多通道输入/输出 (Channels, 如 RGB 图像或特征图)，以及通过堆叠卷积层增大神经元感受野 (Receptive Field) 的范围。
4. **CNN的优势**：通过权重共享和局部连接，CNN能有效利用图像等数据的空间局部性，并且具有一定的平移等变性。相比于MLP，CNN参数更少，更不容易过拟合，训练更高效，特别适合处理网格状数据 (如图像、时间序列)。1D MNIST的例子也证明了CNN在相同任务上比MLP参数少得多，且测试误差更低。

总的来说，这节课重点介绍了CNN的基本原理、结构 (卷积层、权重共享、多通道、感受野等) 及其相比于传统全连接网络在处理图像等空间数据时的优势 (参数效率高、利用局部性、减少过拟合)。

2024.12.11 11.00-14.00

最后一节实验，先下载了贼大的数据集。然后这个实验是关于应用DL解决真实世界的CV问题：使用深度学习识别野外相机陷阱捕获的动物物种。实验要求实现数据加载功能 (CTDataset类) 和训练评估函数 (train和validate)，并比较从头训练和迁移学习两种方法的效果，还探索了不同数据集大小和分割方式对模型性能的影响。但总体不难写。

2024.12.12 9.00-10.00

最后一课了xs, 先把上一节的CNN讲完了, 然后新开了一章讲了几个示例 (Regularization: Ladder; Zero-shot: CLIP; 3D scan: heightfields), 最后总结了一下 (还推荐我们学transformer和ViT, 笑死)。鼓掌!!!

[25 EmbeddingsAndReconstructionExamples.pptx](#)

最后一节课的内容可以总结为以下几点:

1. **补充CNN知识与架构**: 继续介绍了CNN中除卷积层外的其他重要组件, 如下采样 (如最大池化 Max Pooling)、上采样 (如转置卷积 Transpose Convolution) 以及用于调整通道数的1x1卷积。回顾了经典CNN架构 (如AlexNet, VGGNet) 及其在ImageNet等大规模数据集上的成功, 也指出了深度增加带来的参数量和过拟合挑战。
2. **高级学习范式与损失函数**: 简要介绍了利用部分有标签和大量无标签数据的半监督学习方法 (如 Ladder Network), 以及用于学习相对关系的损失函数 (如Triplet Loss, 常用于人脸识别、图像检索等, 目的是让同类样本距离近, 异类样本距离远)。
3. **迁移学习与零样本学习 (CLIP)**: 讨论了将在一个任务上预训练好的模型应用于新任务 (迁移学习/微调 Fine-tuning) 的常用策略 (如冻结部分层)。重点介绍了OpenAI的CLIP模型, 它通过在海量的“图像-文本”配对数据上进行对比学习, 构建了一个统一的图文语义空间。这使得模型能够理解图像内容与文本描述之间的关联, 从而实现零样本学习 (Zero-Shot Learning) ——即不需要针对新的下游任务进行微调, 仅通过提供类别名称 (文本描述), 就能让模型在新图像上进行分类预测。
4. **课程总结与实践建议**: 最后进行了总结, 强调了神经网络实践中的关键点: 模型存在多参数和局部最优问题, 需要正则化 (如数据增强、早停等); 数据预处理至关重要; 需仔细选择损失函数; 要学会高效迭代和调试 (如观察训练曲线); 并鼓励从满足需求的最简单模型开始。

总的来说, 这节课在完成CNN介绍的基础上, 进一步拓展到了更高级的学习范式 (半监督、零样本学习), 特别是介绍了强大的多模态模型CLIP, 并给出了贯穿整个课程的实践性总结和建议。