

机器学习入门：线性判别分析 (LDA)

线性判别分析 (LDA) 是一种经典的机器学习算法，用于在低维空间中最大化类间方差并最小化类内方差。它广泛应用于模式识别、人脸识别和文本分类等领域。

本文将介绍 LDA 的基本原理、数学推导以及其在实际应用中的使用。

LDA 的核心思想是通过寻找最佳的投影方向，使得不同类别的数据在投影后的空间中尽可能分离。假设我们有 n 个样本，每个样本属于 K 个类别之一。我们将这些样本投影到一个 d 维空间中，使得类间方差最大化，同时类内方差最小化。

在 LDA 中，我们定义类间方差 S_B 和类内方差 S_W 。目标是最大化 S_B 并最小化 S_W 。这可以通过求解广义特征值问题来实现。具体来说，我们需要找到最佳的投影方向 w ，使得 $w^T S_B w / w^T S_W w$ 最大化。

在实际应用中，LDA 通常用于降维和分类。通过找到最佳的投影方向，我们可以将高维数据投影到低维空间中，从而减少计算复杂度并提高模型的泛化能力。LDA 在许多领域都有广泛的应用，如人脸识别、文本分类和生物信息学。

总结来说，LDA 是一种有效的降维和分类算法。通过最大化类间方差并最小化类内方差，LDA 能够在低维空间中保持数据的可分性。希望本文能帮助你更好地理解 LDA 的原理和应用。

GDA 模型中，我们假设数据服从高斯分布，且不同类别的数据具有不同的均值和协方差。通过最大化似然函数，我们可以推导出最优的决策边界。对于线性可分的数据，GDA 的决策边界是线性的；对于非线性可分的数据，GDA 的决策边界是非线性的。

在 GDA 中，我们假设数据服从高斯分布，且不同类别的数据具有不同的均值和协方差。通过最大化似然函数，我们可以推导出最优的决策边界。对于线性可分的数据，GDA 的决策边界是线性的；对于非线性可分的数据，GDA 的决策边界是非线性的。