**机器学习纸质课程作业**

1. **证明题**
   1. **证明是凸函数，即对，都有**

**证明：**

**目标不等式**：  
我们需要证明

两边同时乘以 2，等价于

**展开左边项**：  
令 ，则

其中  表示点积。展开点积：

**计算差值**：

代入展开式：

简化：

注意  和 ，所以：

括号中的项为 ：

**分析差值**：  
由于 ，有 ，且 （范数平方非负），因此

这意味着

即

**回到函数**：  
将上述不等式乘以 ，得到

也就是

**结论：**

因此，函数  满足凸函数的定义，是凸函数。

* 1. **证明两个凸函数的和仍然是凸函数，因此支持向量机的目标函数**

是凸函数。

**证明一：两个凸函数之和是凸的**

设 与 在同一凸集 上凸。取任意 与任意 。由凸性分别有

对两式逐项相加得到

即

因此 在 上凸。

**证明二：正标量乘保凸性**

若 凸且 ，则对所有 ：

所以 也凸（若 则会反向，不再保持凸性）。

综上

是凸函数。

* 1. **请证明： 是一个凹函数,即对任意，，有：**

**（提示，请先证明：）**

**应用该定理可以直接得到支持向量机对偶问题的目标函数**

**是凹函数，因此最大化此目标函数有唯一解。**

**1) 先证明提示的不等式**

令 与 为定义在同一集合 上的任意实值函数。我们要证

证明：对任意固定的 ，都有  
且 。因此

对所有 求下确界时，左边的下确界不会小于右边的常数，即

证毕。

**2) 证明 是凹函数**

回忆拉格朗日函数的形式（对任意固定 ）：

对于固定的 ， 是关于 的**仿射函数**（即线性加常数）。仿射函数既是凸的也是凹的；对任意两组乘子 与 和任意 ，

因此

用第 1) 节的推广结论（对带系数的线性组合取 inf 的不等式），我们得到

也就是

这正是 关于 的**凹性**定义。证毕。

**3) 应用到支持向量机对偶问题**

在支持向量机的对偶问题中，目标函数为：

这个函数可以看作是一个特殊的θ(α)函数（没有β参数），其中：

* 线性项对应f(ω)部分
* 二次型对应约束部分

由于：

1. 线性函数是凸的也是凹的;
2. 二次型是凹函数（因为核矩阵是半正定的，其负二次型是凹的）;
3. 凹函数的和仍然是凹函数。

因此，整个目标函数是凹函数。

由于目标函数是凹函数且约束条件是线性的，这个最大化问题是一个凸优化问题，有唯一全局最优解。

**结论：** 我们证明了θ(α,β)是凹函数，并应用这个结论说明了支持向量机对偶问题的目标函数是凹函数，因此最大化此目标函数有唯一解。

* 1. **选做题**

**(a) 在二维特征空间中不存在能被线性分类器“打散”的 4 个点**

**证明**：

1. 我们先排除退化情形（有三点共线等）——若有三点共线，更容易找出不可分的标记（例如把共线的一端标为 +，另一端标为 −，中间点同一侧等），因此只需论证一般位置（无三点共线）的情况即可；随机取点时一般位置的情况以概率 1 发生。
2. 对无三点共线的四点，凸包要么是一个四边形（四点位于凸位置，即四个顶点按环形排列），要么是一个三角形（其中一个点落在其余三点构成的三角形内部）。
   * 若凸包是三角形（即有一个点在其它三个点的凸包内部），记内部点为 ，其余三个为 。把 标为 +1，而把 都标为 −1（三者同号），任何将 与 分开的半空间（由直线给出）都必须把包含 的半空间与包含 的半空间分开。但因为 在 的凸包内部，任何包含 的凸半空间的凸包必然也包含 因而不可能有单一的半空间把 与 分开。由此该标记不可分。
   * 若凸包是凸四边形（四点按环形排列无内点），将四点按环形编号为 。考虑“交替标记”：给 标为 +1，给 标为 −1（或者相反）。任何半空间与凸四边形交集的顶点集合在环形上必是**连通的区间**（半空间切割多边形后留下的顶点是相邻的一段），因此不可能得到交替出现的模式（+、−、+、−）。更形式地：对任意直线，四边形顶点被分到直线两侧时，同一侧的顶点在原多边形循环序列上形成一个连块；交替标记要求同号顶点不成连块，故无法用一条直线实现。 因此交替标记不可分。

两个情况已覆盖所有无退化的四点布置，故对任意 4 点总能找到不可分的标签组合，因此平面线性分类器不能“打散”任意 4 点。证毕。

（因此平面线性分类器的 VC 维为 3——存在某些 3 点可被打散，且不存在任意 4 点被打散。）

**(b) 在线性分类器（超平面 ）下， 的 VC 维为**

要证两件事：  
(i) 存在 个点能被线性分类器打散；  
(ii) 任意 个点不能被线性分类器打散（即 VC ≤ ）。可得 VC = 。

**(i) 构造能被打散的 个点**

取任意 个**仿射无关**的点 。仿射无关意味着这些点的增广矩阵

是一个 的可逆矩阵。现在给任意二分类标签 （共 种），我们要找 使得

更强地，我们可以要求确切地拟合实数值（不仅仅是符号），例如解线性方程组

这是 个线性方程、 个未知数（ 的 个分量和标量 ），系数矩阵就是上面的增广矩阵的转置，因仿射无关而可逆，因此方程组有唯一解。求得的 使得 （严格等于 ±1），于是符号完全匹配任意给定的 。因此任意标签都能被正确分类，说明这 个点被打散。故 VC ≥ 。

**(ii) 任意 个点不能被打散（VC ≤ ）**

任意 个点在 中必有仿射依赖（因为维数为 ，任何 个点线性/仿射相关）。由 **Radon 定理**：任意 个点可以被划分为两组 与 （互不相交），使得这两组点的凸包相交，即 。

现在取这种 Radon 划分，并把 的点全部标为 +1，把 的点全部标为 −1。若存在一个超平面把 +1 与 −1 完全分开，则正类（+1 点）的半空间是一个凸集，且包含 的所有点；因此它也应包含 。类似，负类的半空间包含 。但因为这两凸包有交点，任何将两个凸包严格分开的超平面都不存在（两个凸包相交表示不能通过超平面把它们分离成不同的开半空间）。因此该标记不可被超平面分离。由此任意 个点总存在一种标记不可分，故不能被打散，得 VC ≤ 。

结合 (i)、(ii) 得 VC = 。证毕。

**(c) 由 (b) 推出**

明确题意：

* 在固定的样本数 下，随机从 中取 个点（例如从连续分布独立同分布抽取），并对这 个点随机标记 +1/−1（独立均匀），问：随着维度 增大，点集被线性分类器线性可分（即存在一个超平面把正负两类完全分开）的概率 的极限。

**证明要点**：

* 若 ，并且点为一般位置（概率 1 情形：从连续分布随机取点，任意 个点几乎肯定仿射无关），则这些点仿射无关，从 (b) 我们知道对任意标签都存在一个超平面精确拟合（解线性方程组），因而任意标签组合均线性可分 → **对于这种点的具体位置，任意标签下都可分**。因此当 时，随机取的 个点几乎必是仿射无关，从而**对于任意标签**都线性可分，故此时 （严格来说为 1 在概率意义上：除去测度为 0 的退化点集）。
* 因此当 （固定 ），存在某个阈值 使得对所有 有 。所以

更直观地：当维度远大于样本数时，样本点几乎总是处于仿射无关的位置（没有线性/仿射关系），从而在 维空间中总能找到一个超平面来实现任意的分配；因此随机标记几乎总被线性分隔。

1. **请写出以下问题的对偶问题，并化成最简形式**
2. **-norm SVM Classification**

原问题：

**对偶问题**：

引入拉格朗日乘子 对应于约束 。

拉格朗日函数：

消去原变量 ：

代回得：

这是 -SVM 的最简对偶形式。

1. **SVM Regression（选做）**

原问题：

**对偶问题**

最大化（Maximize）:

1. **Kernel Ridge Regression**

原问题：

**对偶问题**：

拉格朗日函数，令为约束乘子：

求偏导：

代回：

**无约束对偶形式**（线性方程形式）：

从而 。

1. **Entropy Maximization Problem**

原问题：

**对偶问题**：

拉格朗日函数：

对求导并设为0：

代回原约束 ：

代回目标函数得对偶形式：

1. **包含点集的最小球问题**

**对偶推导**

对第 个不等式引入拉格朗日乘子 ，对 引入 。拉格朗日函数：

驻点条件（对 求偏导并令零）给出

把 代回 L 并消去其他项（ 项为零），得到对偶目标：

1. **对偶问题（最简形式）**

（这是一个带线性约束与简单界的二次凸规划问题。）

1. **核化（取 或一般核）**

将内积替换为核函数 ，对偶写成核矩阵形式：

用向量／矩阵记号（令 为核矩阵，， 为系数向量），目标可写为

1. **用核表示**

若索引 满足 （即对应约束为等号的“支撑点”），则由 KKT 有  
。用 展开，得

因此 的核形式为

（任取一个 使 ；若有多个支撑点，任一支撑点代入均应给出相同的 ）。