hw2-ans

1. 给定约束$x+y+z=9$与$x+2y+3z=20$，采用拉格朗日乘子法最小化$f(x,y,z)=x^2+y^2+z^2$。

   解：let $g_1(x,y,z)=x+y+z-9$, $g_2(x,y,z)=x+2y+3z-20$.

   $L(x,y,z,\lambda_1,\lambda_2)=f(x,y,z)+\lambda_1 g_1(x,y,z)+\lambda_2 g_2(x,y,z)$

   let $\frac{\partial L(x,y,z,\lambda_1,\lambda_2)}{\partial i}=0$, where $i\in\{x,y,z,\lambda_1,\lambda_2\}$, then:

   $$\begin{bmatrix}2&0&0&1&1\\0&2&0&1&2\\0&0&2&1&3\\1&1&1&0&0\\1&2&3&0&0\end{bmatrix}\begin{bmatrix}x\\y\\z\\\lambda_1\\\lambda_2\end{bmatrix}=\begin{bmatrix}0\\0\\0\\9\\20\end{bmatrix}$$

   then:

   $$\begin{bmatrix}x\\y\\z\\\lambda_1\\\lambda_2\end{bmatrix}=\begin{bmatrix}2\\3\\4\\-2\\-2\end{bmatrix}$$

   The minimum is: $f(2,3,4)=29$.

注意⚠️：严格来讲此处应当根据二阶导数判断一次极值类型（未作为扣分点）。然而在有约束情况时，不应直接根据f的Hessian Matrix来判断，而应当根据其Bordered Hessian来判断。

---

2. 已知正样本点$(6,5),(9,8),(9,9),(5,9),(6,6)$与负样本点$(2,1),(0,1),(2,4),(0,3),(1,3)$。

   (1) 若已知支持向量为$(6,5),(5,9),(2,4)$，求w,b。（5分）

   解：According to the definition of support vectors:

   $$\begin{bmatrix}6&5&1\\5&9&1\\2&4&1\end{bmatrix}\begin{bmatrix}w_1\\w_2\\b\end{bmatrix}=\begin{bmatrix}1\\1\\-1\end{bmatrix}$$

   Then $\begin{bmatrix}w_1\\w_2\\b\end{bmatrix}=\begin{bmatrix}0.471\\0.118\\-2.411\end{bmatrix}$

   (2) 试根据上述模型预测新样本点$(1,15),(4,2)$的类别。（5分）

   解：$f(x)=w_1x_1+w_2x_2+b$

   $f((1,15))=-0.1764705882352935<0$

   $f((4,2))=-0.2941176470588234<0$

   (3) 证明：对于任意数据分布，若其线性可分，则正负样本中均存在至少一个支持向量。（10分）

解：假设在负样本中没有支持向量。用k表示正样本中某个支持向量的index。

将优化后的参数表示为w^*, b^*，将分隔超平面与点i的几何间隔表示为d_i，即$d_i = y^{(i)}(w^{*T}x^{(i)} + b^*)/||w^*||$。

当我们固定w^*并对b^*进行轻微扰动$b \to b + \delta b$，其中$\delta b > 0$，那么对于所有正样本i^+，有$d_{i^+} \to d_{i^+} + \frac{\delta b}{||w^*||}$；对于所有负样本$i^-$，有$d_{i^+} \to d_{i^-} - \frac{\delta b}{||w^*||}$。

可以验证，在所有样本中，k仍旧是最靠近新超平面$(w^*, b^* + \delta b)$的样本，因为δb非常小。然而，k的几何间隔增加了，这与(w^*, b^*)是最优参数的假设相矛盾。

对于只有负样本中有支持向量的情况，同理。

因此，负样本和正样本中都应该存在支持向量。

注意⚠️：（3）中更简单的证明方法是通过对偶形式。SVM的对偶形式中的约束条件为$\sum_i \alpha_i y_i = 0$，由于非支持向量的$\alpha_i = 0$，因此自然要求正负样本中均具有支持向量来使该式成立。

---

3. 已知正样本点$(0,1),(-1,-1),(1,-1)$与负样本点$(2,2),(-2,2),(0,-2)$。容易观测到，该数据集线性不可分。现采用RBF kernel（$\sigma = 1$）SVM对其进行分类，并且已知优化得到的α_i依次为$1.01, 1.55, 1.55, 1.02, 1.02, 2.08$。

(1) 计算模型优化得到的b（考虑到精度问题，采用上述样本点中的第一个支持向量作为计算依据）。(10分)

解：由每个$\alpha_i > 0$，可得每个样本都是支持向量。使用第一个支持向量$(x^{(1)}, y^{(1)}) = ((0,1)^T, 1)$来计算b。

$$b^* = y^{(1)} - \sum_{j=1}^{6} \alpha_j y^{(j)} k(x^{(1)}, x^{(j)}) = y^{(1)} - \sum_{j=1}^{6} \alpha_j y^{(j)} e^{-\frac{||x^{(1)} - x^{(j)}||^2}{2}} \tag{1}$$

代入数值$(x^{(j)}, y^{(j)})$ for j = 1,2,3,4,5,6，计算出$b^* = -0.0739$

(2) 写出该模型所对应的非线性分割线**方程**（写出形式即可，无需代入具体数值）(10分）。

解：非线性分割线**方程**：

$$\sum_{j=1}^{6} \alpha_j y_j e^{-\frac{||x - x^{(j)}||^2}{2}} + b = 0 \tag{2}$$

注意⚠️：是非线性分割方程，没有写出该等式的都不给该小题分。

(3) 验证$(2,2)$样本点得到了正确分类，并根据该模型判断$(-2,-2)$的类别（10分）。

解：由$f(x) = \sum_{j=1}^{6} \alpha_j y_j e^{-\frac{||x - x^{(j)}||^2}{2}} + b^*$，$b^* = -0.0739$

代入$(2,2)$，$f((2,2)) = \sum_{j=1}^{6} \alpha_j y_j e^{-\frac{||(2,2)^T - x^{(j)}||^2}{2}} + b^* = -1.000799 < 0$；因此$(2,2)$样本点是负样本。

代入$(-2,-2)$，$f((-2,-2)) = \sum_{j=1}^{6} \alpha_j y_j e^{-\frac{||(-2,-2)^T - x^{(j)}||^2}{2}} + b^* = 0.226432 > 0$

注意⚠️：不写出计算过程和计算结果的会扣除5分，部分同学出错在于b^*忘记加了。

---

4. 已知训练数据集有正例点$x_1 = (4,3), x_2 = (3,3), x_3 = (1.5, 1.5)$，负例点$x_4 = (1,1)$。训练一个硬间隔的SVM，分隔超平面为$2x^{(1)} + 2x^{(2)} - 5 = 0$。

(1) 现重新训练一个带松弛变量的软间隔SVM。假设求得的松弛变量$\xi_1=0,\xi_2=0,\xi_3=1.5,\xi_4=0$,请问加松弛变量前后支持向量分别是哪些，以及新的软间隔SVM的分隔超平面。

解：加松弛变量前：$f(x)=2x^{(1)}+2x^{(2)}-5$，我们可为x_i计算函数间隔$y_if(x_i)$。

$y_1f(x_1)=2*4+2*3-5=9$

$y_2f(x_2)=2*3+2*3-5=7$

$y_3f(x_3)=2*1.5+2*1.5-5=1$

$y_4f(x_4)=-(2*1+2*1-5)=1$

可根据函数间隔为1来得到支持向量为x_3,x_4。**(也可以通过计算几何间隔来得到）【3分】**

加松弛变量后：我们可知$\xi_i\neq 0$的点一定是支持向量，可得x_3为支持向量，又知对线性可分数据，正负样本均至少存在一个支持向量，而数据集中负样本只有一个x_4，因此x_4也为支持向量，进一步可得到优化不等式：

$4w_1+3w_2+b\geq 1-0=1$

$3w_1+3w_2+b\geq 1-0=1$

$1.5w_1+1.5w_2+b=1-1.5=-0.5$（由支持向量$x_3$得）

$-(w_1+w_2+b)=1-0=1$（由支持向量x_4得）

结合等式3、4可计算出$w_1+w_2=1$，$b=-2$，由于我们目标函数最小化的是$\frac{1}{2}w_1^2+\frac{1}{2}w_2^2+C\sum_i\xi_i$，当$w_1=w_2=\frac{1}{2}$时，目标函数最小。因此我们可计算得到$x_2$也是支持向量。

综上所述，加松弛变量后的支持向量由x_2,x_3,x_4。

软间隔SVM的分割超平面为$\frac{1}{2}x^{(1)}+\frac{1}{2}x^{(2)}-2=0$。

注意⚠️：不少同学没有考虑x_3也是支持向量，均扣除3分。

(2) 分别计算加松弛变量前的**硬间隔**和加松弛变量后的**软间隔**，并根据该结果来解释加松弛变量对模型的好处。

间隔是：根据老师上课定义，$\xi=0$的支持向量到超平面的几何距离。

解：加松弛变量前的硬间隔是：$\frac{|wx+b|}{||w||}=\frac{|2*1.5+2*1.5-5|}{2\sqrt{2}}=\frac{1}{2\sqrt{2}}$

加松弛变量后的软间隔是：$\frac{|wx+b|}{||w||}=\frac{|0.5*3+0.5*3-2|}{0.5\sqrt{2}}=\frac{2}{\sqrt{2}}$

加松弛变量增大了间隔大小，减少模型对ourlier的依赖性，使得模型更为鲁棒。

注意⚠️：有同学将距离乘2得到间隔，和老师上课定义的margin不一致，酌情扣了2分。

(3) 根据（1）给定的松弛变量，写出每个数据点的hinge loss，并从hinge loss角度来解释为什么支持向量机能保持解的稀疏性。

解：由$hinge\ loss=max\{0,1-y_i(w^Tx_i+b)\}$得：

$x_1: hinge\ loss=0,$

$x_2: hinge\ loss=0,$

$x_3: hinge\ loss=1.5,$

$x_4: hinge\ loss=0,$

hinge loss = 0的区域对应$y_i(w^T x_i + b) > 1$的点，这些点分类正确且距离超平面较远（>1）对loss不作贡献，而对于$y_i(w^T x_i + b) \leq 1$的点，loss惩罚较高，即尽可能优化的是分类错误或距离平面较近的点，从而保证了解的稀疏性（只有$\alpha > 0$ 的点和超平面的确定有关，即支持向量）。