# 作业（一）

**Q1**，试解释机器学习中过拟合的概念。有哪些方法可以在一定程度上缓解或解决过拟合的问题？（10 分）

**Q2**，针对一个线性回归问题，假设给定一个具有$n$个数据样本的训练数据集，其中对于每个样本$\{x_i, y_i | i = 1,2,\ldots,n\}$，$x_i$是样本特征，$y_i$是其对应的目标值（真值）。

（1）请利用数学公式定义该线性回归模型（模型参数请用$w_i$表示），该模型是利用一条直线去拟合此数据集的所有数据。此外，请写出基于平方误差总和（SSE）的线性回归损失函数（目标损失函数），该目标损失函数的优化目标是试图找一条能使得该目标函数值最小的直线。（7 分）

（2）为了避免过拟合问题，通常在损失函数中增加一个 L2 范数正则化项（即岭回归），请用数学公式定义该完整的损失函数，并请推导参数$W(W = \{w_i\}_{i=0}^{n})$的封闭最优解（求解其梯度，并令梯度为 0，求出参数 W 的封闭解形式）。(13 分)

**Q3**，假设利用梯度下降法优化一个单变量的函数：$\min_{\theta} J(\theta) = \theta^2 + 2$，假设$\hat{\theta}$初始的起点为$\hat{\theta} = (1,3)$, 学习率为$\alpha = 0.1$，试利用梯度下降法优化该函数，求该函数的最小值以及$\hat{\theta}$。（20 分）

**Q4,** 已知正例点$x_1 = (1,2), x_2 = (2,3), x_3 = (3,3)$，负例点$x_4 = (2,1), x_5 = (3,2), x_6 = (3,0)$,试求最大间隔分离超平面和分类决策函数，并在图上画出分离超平面、间隔边界及支持向量。（25 分）

**Q5**，假设在一个二维空间（平面）里给定 5 个点，{2,0}，{3,1},{2,1},{0,5},{-1,5},这些点属于两个聚类（K=2），试利用 K-means 算法找到这些点所属的聚类以及每个聚类的中心点。（聚类的初始化中心点随机产生，在下面写出 K-means 算法的详细步骤，直至收敛）。（25 分）