1. **请默写出拉格朗日乘子法的一般形式。见Page 6中，同时包含等式条件和不等式条件下的拉格朗日乘子法的问题所对应的一般形式，和解法的一般形式。（不是写对SVM的求解过程，只写一般形式，而且写出解法的一般套路即可，“对L求梯度=0”这一步为止）**

$$
\begin{aligned}
&\min_w f(w) \\
\text{s.t.} \qquad & g_k(w) \le 0, \quad k = 1, 2, \cdots, K \\
& h_l(w) = 0, \quad l = 1, 2, \cdots, L
\end{aligned}
$$

根据拉格朗日乘子法，我们可以构造拉格朗日函数

$$
\begin{gathered}
L(w, \alpha, \beta) = f(w) + \sum_{k=1}^{K} \alpha_k g_k(w) + \sum_{l=1}^{L} \beta_l h_l(w) \\
\min_w \max_{\alpha, \beta : \alpha_k \ge 0} L(w, \alpha, \beta)
\end{gathered}
$$

KKT条件如下：

$$
\begin{aligned}
\forall p, \frac{\partial}{\partial w_p} L(w, \alpha, \beta) = 0 \\
\forall l, \frac{\partial}{\partial \beta_l} L(w, \alpha, \beta) = 0 \\
\forall k, \alpha_k g_k(w) = 0 \\
\forall k, g_k(w) \le 0 \\
\forall k, \alpha_k \ge 0
\end{aligned}
$$

2. **Page 8：仅考虑当“数据线性可分”的情况下，从拉格朗日乘子法的角度，解释为什么SVM的参数w的方向仅决定于数据中的支持向量。此问题作答包括三个方面。**
   a. 什么叫“支持向量”。
   b. 如何推导出w的解的形式。
   c. 从拉格朗日乘子法的角度解释一下（不一定严谨的证明），为什么仅有支持向量所对应的alpha_i才可能大于0。或者解释为什么不是支持向量的样本的alpha_i必须等于0。
   综合上面三个方面，可以解释为什么SVM的参数w的方向仅决定于数据中的支持向量。
   - a. 支持向量是指满足 $-y_i(w^\top X_i + b) + 1 = 0$ 条件的训练样本点。
   - b. 在线性可分的情况下，SVM的优化问题可以表示为：

$$
\begin{aligned}
&\min_{w,b} \frac{1}{2} \|w\|^2 \\
\text{s.t.} \qquad & \forall i, -y_i \left(w^\top X_i + b \right) + 1 \le 0
\end{aligned}
$$

     拉格朗日方程如下：

$$
\begin{gathered}
L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i \left(w^\top X_i + b \right) - 1 \right] \\
\min_{w,b} \max_{\alpha : \alpha_i \ge 0} L(w, b, \alpha)
\end{gathered}
$$

考虑 $\frac{\partial L(w,b,\alpha)}{\partial w}=0$，得到

$$w-\sum_{i=1}^{n}\alpha_i y_i X_i=0$$
$$w=\sum_{i=1}^{n}\alpha_i y_i X_i$$

- c. 假设第 i 个样本点不是支持向量，则有

$$y_i(w^\top X_i+b)-1>0$$

在进行 $\max\limits_{\alpha:\alpha_i>0}$ 时，会导致 $\alpha_i=0$。
因此不是支持向量的 α_i 必须等于 0。

3. **Page13：请默写出带outlier时，SVM的目标函数的定义（带不等式条件的定义），以及SVM的等效目标函数的定义（不带不等式条件的定义），以及SVM在拉格朗日乘子法下的损失函数。（只写定义，不写求解过程）**

$$\begin{aligned}&\min_{\xi,w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\xi_i\\ \text{s.t.}\quad & y_i\left(w^\top X_i+b\right)\geq 1-\xi_i,\quad i=1,2,\cdots,n\\ &\xi_i\geq 0,\quad i=1,2,\cdots,n\end{aligned}$$

$$\min_{\xi,w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\max\left(0,1-y_i\left(w^\top X_i+b\right)\right)$$

$$L(w,b,\xi,\alpha,\beta)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\alpha_i\left[y_i\left(w^\top X_i+b\right)-1+\xi_i\right]-\sum_{i=1}^{n}\beta_i\xi_i$$
$$\min_{w,b,\xi:\xi_i\geq 0}\ \max_{\alpha,\beta:\alpha_i\geq 0,\beta_i\geq 0} L(w,b,\xi,\alpha,\beta)$$