1. **Please write the formulation for the least square regression, the ridge regression, the kernel regression, and the LASSO regression.**
least square regression:

$$\|Y - X^\top \beta\|^2$$

ridge regression:

$$\|Y - X^\top \beta\|^2 + \lambda \|\beta\|_2^2$$

kernel regression:

$$\|Y - Kc\|^2 + \lambda c^\top Kc$$

LASSO regression:

$$\|Y - X^\top \beta\|^2 + \lambda |\beta|_1$$

2. **Please write analytic solutions to the least square regression, the ridge regression, the kernel regression, and the spline regression. (不用写证明过程)**
least square regression:

$$\beta = \left(X^\top X\right)^{-1} X^\top Y$$

ridge regression:

$$\beta = \left(X^\top X + \lambda I\right)^{-1} X^\top Y$$

kernel regression:

$$c = (K + \lambda I)^{-1} Y$$

spline regression:

$$\alpha = \left(Z^\top Z + \lambda D\right)^{-1} Z^\top Y$$

3. **Please derive the LDA for two classes.**

$$\begin{gathered}\forall X_i \in \Omega^+, p\left(X_i \mid y = +1\right) \sim \mathrm{N}\left(\mu^+, \mu^-\right) \\ \forall X_i \in \Omega^-, p\left(X_i \mid y = -1\right) \sim \mathrm{N}\left(\mu^-, \mu^+\right)\end{gathered}$$

$$
\begin{aligned}
\sigma_{\text{between}}^2 &= \left[(\mu^+ - \mu^-)^\top \beta\right]^2 \\
\sigma_{\text{within}}^2 &= n_{\text{pos}} \sigma_{\text{pos}}^2 + n_{\text{neg}} \sigma_{\text{neg}}^2 \\
n_{\text{pos}} &= |\Omega^+| \\
n_{\text{neg}} &= |\Omega^-| \\
\sigma_{\text{pos}} &= \beta^\top \Sigma^+ \beta \\
\sigma_{\text{neg}} &= \beta^\top \Sigma^- \beta
\end{aligned}
$$

$$
\begin{aligned}
S &= \frac{\sigma_{\text{between}}^2}{\sigma_{\text{within}}^2} \\
&= \frac{\left[(\mu^+ - \mu^-)^\top \beta\right]^2}{n_{\text{pos}} \sigma_{\text{pos}}^2 + n_{\text{neg}} \sigma_{\text{neg}}^2} \\
&= \frac{\beta^\top S_B \beta}{\beta^\top S_W \beta}
\end{aligned}
$$

其中

$$
\begin{aligned}
S_B &= \left(\mu^+ - \mu^-\right) \left(\mu^+ - \mu^-\right)^\top \\
S_W &= n_{\text{pos}} \Sigma^+ + n_{\text{neg}} \Sigma^-
\end{aligned}
$$

因为 β 模长不改变结果，可令

$$
\beta^\top S_W \beta = 1
$$

$$
\max_\beta \beta^\top S_B \beta \qquad \text{s.t.} \qquad \beta^\top S_W \beta = 1
$$

$$
\begin{aligned}
& L = \beta^\top S_B \beta - \lambda \left(\beta^\top S_W \beta - 1\right) \\
\Rightarrow \quad & \frac{\partial L}{\partial \beta} = 2 S_B \beta - 2\lambda S_W \beta = 0 \\
\Rightarrow \quad & S_B \beta = \lambda S_W \beta \\
\Rightarrow \quad & S_W^{-1} S_B \beta = \lambda \beta
\end{aligned}
$$

β 是 $S_W^{-1} S_B$ 的特征向量。

因为 $S_B \beta$ 与 $\mu^+ - \mu^-$ 方向相同，

$$
\beta \propto S_W^{-1} \left(\mu^+ - \mu^-\right)
$$