SVM Problems Solutions:

1) **Find a linear model that minimizes the training error,**

In here we use a new $\hat{w}' = (\hat{b}, \hat{w})$ and $x_i' = (1, x_i)$ and for convenience. So by lecture note, we can easy to know that we get that the minimum will occur at $\hat{w}$ that satisfy

$$\left[\begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_m \end{bmatrix}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}\right]\hat{w}' = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_m \end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

So, let $\Sigma = \left[\begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_m \end{bmatrix}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}\right] = \begin{bmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m {x_i}^2 \end{bmatrix}$

If Σ is invertible, $\mathrm{m}\sum_{i=1}^m {x_i}^2 - (\sum_{i=1}^m x_i)^2$

$$\Sigma^{-1} = \begin{bmatrix} \frac{\sum_{i=1}^m {x_i}^2}{\mathrm{m}\sum_{i=1}^m {x_i}^2 - \left(\sum_{i=1}^m x_i\right)^2} & \frac{-\sum_{i=1}^m x_i}{\mathrm{m}\sum_{i=1}^m {x_i}^2 - \left(\sum_{i=1}^m x_i\right)^2} \\ \frac{-\sum_{i=1}^m x_i}{\mathrm{m}\sum_{i=1}^m {x_i}^2 - \left(\sum_{i=1}^m x_i\right)^2} & \frac{m}{\mathrm{m}\sum_{i=1}^m {x_i}^2 - \left(\sum_{i=1}^m x_i\right)^2} \end{bmatrix}$$

$$\hat{w}' = \Sigma^{-1}\left[\sum_{i=1}^m y_i \quad \sum_{i=1}^m y_i x_i\right]^T$$

If Σ is not invertible or poorly conditioned, we can assume a $\lambda > 0$ such that $\Sigma' = \Sigma + \lambda I$,

$$\hat{w}' = \Sigma'^{-1}\left[\sum_{i=1}^m y_i \quad \sum_{i=1}^m y_i x_i\right]$$

2) **Assume there is some true linear model, such that...**

From lecture we know that since $\epsilon_i \sim N(0, \sigma^2)$, we have $\underline{\epsilon} \sim N(0, I\sigma^2)$, so we can get

$$\hat{w}'' \sim \hat{w}' + N\left(\underline{0}, \Sigma^{-1}\sigma^2\right)$$

That is

$$\hat{w}'' \sim N(\hat{w}', \Sigma^{-1}\sigma^2)$$

So we can know that the expect of $\hat{w}''$ is $\hat{w}'$ which means $\mathbb{E}[\hat{w}] = w$ and $\mathbb{E}[\hat{b}] = b$.

So we can know that

$$\mathrm{Var}[\hat{w}] = \mathbb{E}\left[(\hat{w} - \mathbb{E}[\hat{w}])^2\right] = \mathrm{Cov}(\hat{w}, \hat{w}) = {\Sigma^{-1}}_{22}\sigma^2 = \frac{m\sigma^2}{\mathrm{m}\sum_{i=1}^m {x_i}^2 - \left(\sum_{i=1}^m x_i\right)^2}$$

$$\mathrm{Var}[\hat{b}] = \mathbb{E}\left[\left(\hat{b} - \mathbb{E}[\hat{b}]\right)^2\right] = \mathrm{Cov}(\hat{b}, \hat{b}) = {\Sigma^{-1}}_{11}\sigma^2 = \frac{\sigma^2 \sum_{i=1}^m {x_i}^2}{\mathrm{m}\sum_{i=1}^m {x_i}^2 - \left(\sum_{i=1}^m x_i\right)^2}$$

3) **Assume that each x value was sampled from some underlying distribution with expectation…**

From 2). We get

$$\mathrm{Var}[\widehat{w}] = \frac{m\sigma^2}{\mathrm{m}\sum_{i=1}^{m} {x_i}^2 - \left(\sum_{i=1}^{m} x_i\right)^2}$$

$$\mathrm{Var}[\hat{b}] = \frac{\sigma^2 \sum_{i=1}^{m} {x_i}^2}{\mathrm{m}\sum_{i=1}^{m} {x_i}^2 - \left(\sum_{i=1}^{m} x_i\right)^2}$$

And $\mathbb{E}[\mathrm{x}^2] = \frac{1}{m}\sum_{i=1}^{m} {x_i}^2$, $\mathbb{E}[\mathrm{x}]^2 = \frac{1}{m^2}(\sum_{i=1}^{m} x_i)^2$, $\mathrm{Var}[x] = \mathbb{E}[\mathrm{x}^2] - \mathbb{E}[\mathrm{x}]^2$.

So we have:

$$\mathrm{Var}[\widehat{w}] = \frac{\sigma^2}{m}\frac{1}{\frac{1}{m}\sum_{i=1}^{m} {x_i}^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m} x_i\right)^2} = \frac{\sigma^2}{m}\frac{1}{\mathrm{Var}[x]}$$

$$\mathrm{Var}[\hat{b}] = \frac{\sigma^2}{m}\frac{\frac{1}{m}\sum_{i=1}^{m} {x_i}^2}{\frac{1}{m}\sum_{i=1}^{m} {x_i}^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m} x_i\right)^2} = \frac{\sigma^2}{m}\frac{\mathbb{E}[\mathrm{x}^2]}{\mathrm{Var}[x]}$$

4) **Argue that recentering the data…**

For ${x_i}' = x_i - \mathbb{E}[\mathrm{x}]$, we can have

$$\mathbb{E}[\mathrm{x}'] = \mathbb{E}[\mathrm{x} - \mathbb{E}[\mathrm{x}]] = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mathbb{E}[\mathrm{x}])$$

$$\mathbb{E}\left[\mathrm{x}'^2\right] = \mathbb{E}[(\mathrm{x} - \mathbb{E}[\mathrm{x}])^2] = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mathbb{E}[\mathrm{x}])^2 = \mathbb{E}[\mathrm{x}^2] - \mathbb{E}[\mathrm{x}]^2 = \mathrm{Var}[x]$$

For w and b after re-center, we have new variance(error):

$$\mathrm{Var}'[\widehat{w}] = \frac{\sigma^2}{m}\frac{1}{\mathrm{Var}[\mathrm{x}']} = \frac{\sigma^2}{m}\frac{1}{\mathrm{Var}[\mathrm{x}]} = \mathrm{Var}[\widehat{w}]$$

$$\mathrm{Var}'[\hat{b}] = \frac{\sigma^2}{m}\frac{\mathbb{E}\left[\mathrm{x}'^2\right]}{\mathrm{Var}[\mathrm{x}']} = \frac{\sigma^2}{m}\frac{\mathrm{Var}[x]}{\mathrm{Var}[x]} = \frac{\sigma^2}{m}$$

And $\mathrm{Var}[x] = \mathbb{E}[\mathrm{x}^2] - \mathbb{E}[\mathrm{x}]^2$ so $\mathrm{Var}[x] \le \mathbb{E}[\mathrm{x}^2]$, $\frac{\mathbb{E}[\mathrm{x}^2]}{\mathrm{Var}[x]} \ge 1$.

So,

$$\mathrm{Var}'[\hat{b}] \le \mathrm{Var}[\hat{b}]$$

Which means the error on b is minimized.

5)

	$\hat{b}$	$\hat{w}$	$\hat{b}'$	$\hat{w}'$
Expect	5.0736164989159445	0.9992722891965569	106.00011770778859	0.9992722891966043
Variance	14.097971581857914	0.0013819171525701675	0.0004828191987593426	0.0013819171525236787

It obvious make sense.

6)

It's because it just the translation transformation but not change the relative position in the axis. So the slope is not change.

7)

$$\Sigma' = \begin{bmatrix} 1 & \mathbb{E}[x'] \\ \mathbb{E}[x'] & \mathbb{E}\left[x'^2\right] \end{bmatrix}$$

And

$$\mathbb{E}[x'] = \mathbb{E}[x - \mu] = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mathbb{E}[x]) = \frac{1}{m}\sum_{i=1}^{m} x_i - \mathbb{E}[x] = \mathbb{E}[x] - \mathbb{E}[x] = 0$$

$$\mathbb{E}\left[x'^2\right] = \mathbb{E}[(x - \mathbb{E}[x])^2] = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mathbb{E}[x])^2 = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \mathrm{Var}[x]$$

So,

$$\Sigma' = \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{Var}[x] \end{bmatrix}$$

So, by

$$\begin{vmatrix} 1-\lambda & 0 \\ 0 & \mathrm{Var}[x] - \lambda \end{vmatrix} = 0$$

We know

$$\lambda_1' = 1, \lambda_2' = \mathrm{Var}[x]$$

In the same way, we know λ of Σ is: we set

$$a = \mathbb{E}[x], b = \mathbb{E}[x^2]$$

$$\lambda_1 = \frac{1 + b + \sqrt{(b-1)^2 + 4a^2}}{2}, \lambda_2 = \frac{1 + b - \sqrt{(b-1)^2 + 4a^2}}{2}, \lambda_1 > \lambda_2$$

So,. If $\mathrm{Var}[x] > 1$

$$\kappa' = \frac{\mathrm{Var}[x]}{1}, \kappa = \frac{\lambda_1}{\lambda_2}$$

So,

$$\kappa' - \kappa = \frac{(a^2 - b - 1)\left(1 - b + \sqrt{(b-1)^2 + 4a^2}\right)}{1 + b - \sqrt{(b-1)^2 + 4a^2}}$$

$$a^2 - b - 1 = -\mathrm{Var}[x] - 1 < 0$$

$$\sqrt{(b-1)^2 + 4a^2} \geq \sqrt{(b-1)^2} = |b - 1|$$

And

$$b = \mathbb{E}[\mathrm{x}^2] \geq \mathrm{Var}[x] > 1$$

So,

$$\left(1 - b + \sqrt{(b-1)^2 + 4a^2}\right) \geq 1 - b + b - 1 \geq 0$$

So,

$$\kappa' - \kappa = \frac{(a^2 - b - 1)\left(1 - b + \sqrt{(b-1)^2 + 4a^2}\right)}{1 + b - \sqrt{(b-1)^2 + 4a^2}} \leq 0$$

$$\kappa' \leq \kappa$$

The same to $\mathrm{Var}[x] < 1$