Task 2
Not always. It exist when inverse of $X^T \cdot X$ exist.

Task 3`

A. OLS provides the Best Linear Unbiased Estimator that means that if you take any other unbiased estimator, it is bound to have a higher variance then the OLS solution.

Regularization is to add some bias in turn to try to reduce the variance because prediction error, it is a combination of three things:

$$E[(y - \hat{f}(x))^2] = Bias[\hat{f}(x))]^2 + Var[\hat{f}(x)] + \sigma^2$$

The last part is the irreducible error, so we have no control over that. Using the OLS solution the bias term is zero. But it might be that the second term is large. It might be a good idea, (if we want good predictions), to add in some bias and hopefully reduce the variance.

$Var[\hat{f}(x)]$ is the variance introduced in the estimates for the parameters in our model. The linear model has the form

$$y = X\beta + \epsilon, \qquad \epsilon \sim N(0, \sigma^2 I)$$

To obtain the OLS solution we solve the minimization problem

$$arg\ min_{\beta}\ \| y - X\beta \|^2$$

This provides the solution

$$\widehat{\beta}_{OLS} = (XTX)^{-1} X^T y$$

The minimization problem for ridge regression is similar:

$$argmin\beta \| y - X\beta \|^2 + \lambda \| \beta \|^2, \qquad \lambda > 0$$

Now the solution becomes

$$\widehat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T y$$

So we are adding this λI (called the ridge) on the diagonal of the matrix that we invert. The effect this has on the matrix XTX is that it "pulls" the determinant of the matrix away from zero. Thus when we invert it, we do not get huge eigenvalues. But that leads variance of the parameter estimates becomes lower.

B. It can be recast as

$$\sum_{i=1}^{n} (y_i - x_j^T \beta)^2 + \lambda \sum_{i=1}^{p} \beta_j^2$$

$$\sum_{i=1}^{n} (y_i - x_i \beta)^2 + \sum_{i=1}^{p} (0 - \sqrt{\lambda}\beta_j)^2$$

In other words a diagonal matrix whose value is $\sqrt{\lambda}$ and all y are 0

$$\mathbf{Z}_\lambda = \begin{pmatrix} z_{1,1} & z_{1,2} & z_{1,3} & \cdots & z_{1,p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{n,1} & z_{n,2} & z_{n,3} & \cdots & z_{n,p} \\ \sqrt{\lambda} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda} & 0 & \cdots & 0 \\ 0 & 0 & \sqrt{\lambda} & \ddots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sqrt{\lambda} \end{pmatrix} ; \mathbf{y}_\lambda = \begin{pmatrix} y_1 \\ \vdots \\ y_n \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$