1a) Skal vise at

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

Den deriverte kostfunksjonen

$$C(X,\beta) = \frac{1}{n}\|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2$$
$$= \frac{1}{n}(y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta$$

mtp. β^T blir

$$\frac{\partial C(X,\beta)}{\partial \beta^T} = \frac{1}{n}\frac{\partial}{\partial \beta^T}\left((y - X\beta)^T(y - X\beta)\right) + \lambda\frac{\partial}{\partial \beta^T}\beta^T\beta$$
$$= \frac{1}{n}\frac{\partial}{\partial \beta^T}\left((y^T - \beta^T X^T)(y - X\beta)\right) + \lambda\frac{\partial}{\partial \beta^T}\beta^T\beta$$
$$= \frac{1}{n}\frac{\partial}{\partial \beta^T}\left(y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta\right)$$
$$+ \lambda\frac{\partial}{\partial \beta^T}\beta^T\beta$$
$$= \frac{1}{n}(-X^T y + X^T X\beta) + \lambda\beta = 0$$

$\frac{1}{n}$ forsvinner

$$\Rightarrow X^T X\beta + \lambda\beta = X^T y$$

Vet at for en vektor $\vec{a}$, så er
$\vec{a} = I\vec{a}$.

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

b) Skal vise at

$$\tilde{y}_{OLS} = X\beta = \sum_{i=0}^{p-1} u_i u_i^T y$$

ved SVD.

$$X = U\Sigma V^T \Rightarrow X^T = V\Sigma^T U^T$$
$$\beta = (X^T X)^{-1} X^T y$$
$$\Rightarrow \tilde{y}_{OLS} = U\Sigma V^T (X^T X)^{-1} X^T y$$
$$= U\Sigma V^T (V\Sigma^T \underbrace{U^T U}_{I} \Sigma V^T)^{-1} V\Sigma^T U^T y$$
$$= U\Sigma V^T (V\Sigma^2 V^T)^{-1} V\Sigma^T U^T y$$

Vet at V, V^T og Σ^2 kvadratiske og inverterbare:
For to matriser A og B har vi at
$(AB)^{-1} = B^{-1}A^{-1}$

Ergo:
$(V\Sigma^T\Sigma V^T)^{-1} = (V^T)^{-1}(\Sigma^T\Sigma)^{-1}V^{-1}$
hvor $(V^T)^{-1} = V$ og $V^{-1} = V^T$

$$\Rightarrow U\Sigma \underbrace{V^T V}_{I} (\Sigma^2)^{-1} \underbrace{V^T V}_{I} \Sigma^T U^T y$$

$$= U\Sigma(\Sigma^2)^{-1}\Sigma^T U^T y$$
$$= UU^T y = \sum_{i=0}^{p-1} u_i u_i^T y$$

Nå, for Ridge-regresjon:

$$\tilde{y}_{Ridge} = X\beta_{Ridge} = \sum_{i=0}^{p-1} u_i u_i^T \frac{\sigma_i^2}{\sigma_i^2 + \lambda} y$$

hvor vi introduserer hyperparameteren λ.

$$\tilde{y}_{Ridge} = U\Sigma V^T (V\Sigma^T \underbrace{U^T U}_{I} \Sigma V^T + \lambda I)^{-1} V\Sigma^T U^T y$$
$$= U\Sigma V^T (V\Sigma^2 V^T + \lambda I)^{-1} V\Sigma^T U^T y$$

Kan inkludere V og V^T rundt λI fordi $VV^T = I$ og $IV = VI$

$$\Rightarrow U\Sigma V^T (V\Sigma^2 V^T + V\lambda I V^T)^{-1} V\Sigma^T U^T y$$
$$= U\Sigma V^T (V(\Sigma^2 + \lambda I)V^T)^{-1} V\Sigma^T U^T y$$

Husk at $(VV^T)^{-1} = V^{-T}V^{-1} = VV^{-1}$.

$$= U\Sigma \underbrace{V^T V}_{I} (\Sigma^2 + \lambda I)^{-1} \underbrace{V^{-1} V}_{I} \Sigma^T U^T y$$
$$= U\Sigma(\Sigma^2 + \lambda I)^{-1}\Sigma^T U^T y$$
$$= U\Sigma^2(\Sigma^2 + \lambda I)^{-1} U^T y$$

Siden $(\Sigma^2 + \lambda I)$ er en diagonal matrise med ingen 0'er langs diagonalen, er den inverterbar. Det gjelder også produktet $\Sigma^2(\Sigma^2 + \lambda I)^{-1}$, så det kommuterer med U^T.

$$\Rightarrow \tilde{y}_{Ridge} = UU^T\Sigma^2(\Sigma^2 + \lambda I)^{-1} y$$
$$= \underline{\underline{\sum_{i=0}^{n-1} u_i u_i^T \frac{\sigma_i^2}{\sigma_i^2 + \lambda} y}}$$