Derivation of effective degrees of freedom (3.50) on P.68.

Previously in linear regression we met the 'hat matrix' $H = X(X^T X)^{-1} X^T$ s.t $\hat{y} = Hy$

When $\text{rank}(X) = P$, we have that:

$$
\begin{aligned}
\text{trace}(H) &= \text{trace}(X(X^T X)^{-1} X^T) \\
&= \text{trace}(X^T X (X^T X)^{-1}) \quad \text{(commutivity of tr)} \\
&= \text{trace}(\mathbb{I}_P) \\
&= P
\end{aligned}
$$

and this is known as the degrees of freedom used in the model. We would now like similar notation for ridge regression $\rightarrow$ effective degrees of freedom.

$$
\begin{aligned}
\hat{y}^{\text{ridge}} &= X\hat{\beta}^{\text{ridge}} \\
&= X(X^T X + \lambda \mathbb{I})^{-1} X^T y
\end{aligned}
$$

and using my previous derivation of 3.47 we obtain:

$$\ldots$$

$$= \underbrace{UD(D^2 + \lambda\mathbb{I})^{-1} D U^T}_{H^{\text{ridge}}} y$$

Now,

$$
\begin{aligned}
&\text{trace}(H^{\text{ridge}}) \\
&= \text{trace}(UD(D^2 + \lambda\mathbb{I})^{-1} D U^T) \\
&= \text{trace}(D(D^2 + \lambda\mathbb{I})^{-1} D) \quad \text{(commutivity of orthogonal } U\text{)} \\
&= \sum_{j=1}^{P} \frac{d_j^2}{d_j^2 + \lambda} \Rightarrow \text{The effective degrees of freedom}
\end{aligned}
$$