№2

1). $E_\theta(\hat\theta-\theta)^T(\hat\theta-\theta) = \Big/\!\Big/\ \hat\theta = (X^TX-\lambda\mathbb{1})^{-1}X^TY \quad \lambda\in\mathbb{R}.$
$Y = X\theta+\varepsilon,\ E_\theta\varepsilon=0.\ \Big/\!\Big/$

$$= E_\theta(\hat\theta^T\hat\theta - \hat\theta^T\theta - \theta^T\hat\theta + \theta^T\theta) = E_\theta(\hat\theta^T\hat\theta - \hat\theta^T\theta - \theta^T\hat\theta + \theta^T\theta) =$$

$$= E_\theta(\hat\theta^T\hat\theta) - E_\theta(\hat\theta^T\theta) - E_\theta\hat\theta^T\hat\theta + E_\theta\theta^T\theta = E_\theta(Y^TX(X^TX-\lambda)^{-1}(X^TX-\lambda)^{-1}X^TY) -$$

$$- E_\theta(\underbrace{Y^T}_{\theta^TX^T+\varepsilon^T}X(X^TX-\lambda)^{-1}\theta) - E_\theta(\theta^T(X^TX-\lambda)^{-1}X^T\underbrace{Y}_{X\theta+\varepsilon}) + \theta^T\theta =$$

$$= E_\theta\big[(\theta^TX^T+\varepsilon^T)X(X^TX-\lambda)^{-1}(X^TX-\lambda)^{-1}X^T(X\theta+\varepsilon) - \theta^TX^TX(X^TX-\lambda)^{-1}\theta - \theta^T(X^TX-\lambda)^{-1}X^TX\theta$$

$$+ \theta^T\theta = \theta^TX^TX(X^TX-\lambda)^{-1}(X^TX-\lambda)^{-1}X^TX\theta + E_\theta[\varepsilon^TX(X^TX-\lambda)^{-1}(X^TX-\lambda)^{-1}X^T\varepsilon] -$$

$$- \theta^T[X^TX(X^TX-\lambda)^{-1} + (X^TX-\lambda)^{-1}X^TX - 1]\theta = \theta^T[X^TX(X^TX-\lambda)^{-1}(X^TX-\lambda)^{-1}X^TX -$$

$$- X^TX(X^TX-\lambda)^{-1} - (X^TX-\lambda)^{-1}X^TX + 1]\theta + E_\theta[\varepsilon^T\underbrace{X(X^TX-\lambda)^{-1}(X^TX-\lambda)^{-1}X^T}_{\Gamma}\varepsilon]$$

Для вычисления слагаемого $E_\theta[\varepsilon^T\Gamma\varepsilon]$ добавим доп условий на ε: Пусть компоненты вектора ε независимы $\Rightarrow$ $\cancel{E_\theta[\varepsilon^T\Gamma\varepsilon] = E_\theta\varepsilon^i\Gamma_{ij}\varepsilon_j}$

$$\Rightarrow E_\theta[\varepsilon^T\Gamma\varepsilon] = \sum_{i,j}E_\theta(\varepsilon^i\Gamma_{ij}\varepsilon_j) = \Big/\!\Big/ \text{из-за независимости и } E\varepsilon=0 \text{ останутся лишь слагаемые с } i=j \Big/\!\Big/ =$$

$$= tr\,\Gamma\ E_\theta(\varepsilon^T\varepsilon) = tr\,\Gamma\ E_\theta(\varepsilon_1^2+\varepsilon_2^2+\dots+\varepsilon_n^2) = \Big/\!\Big/\ D\varepsilon = E(\varepsilon^2) = E(\varepsilon^T\varepsilon)\ \Big/\!\Big/ = tr\,\Gamma\,D\varepsilon. =$$

$$= tr(X(X^TX-\lambda)^{-1}(X^TX-\lambda)^{-1}X^T)\,D\varepsilon.$$

Собирая всё вместе получим ответ:

$$E_\theta(\hat{\Theta}-\Theta)^T(\hat{\Theta}-\Theta) = \Theta^T[X^TX(X^TX-\lambda)^{-1}(X^TX-\lambda)^{-1}X^TX - X^TX(X^TX-\lambda)^{-1} - (X^TX-\lambda)^{-1}X^TX +$$

$$+1] + tr\left(X(X^TX-\lambda)^{-1}(X^TX-\lambda)^{-1}X^T\right)\cdot D\varepsilon.$$

2). Покажем, что $\hat{Y} = X\Theta$ не $\perp$ $e = Y - \hat{Y}$ в ридж-регресии

Покажем это геометрически:

Задача регресии имеет решение $\hat{\Theta} = \underset{\Theta \in \mathbb{R}^d}{\arg\min} \|Y - X\Theta\|_2^2$, т.е. $dist(Y, \mathcal{L}(X))$, где $\mathcal{L}(X) = \{X\Theta \mid \Theta \in \mathbb{R}^d\}$

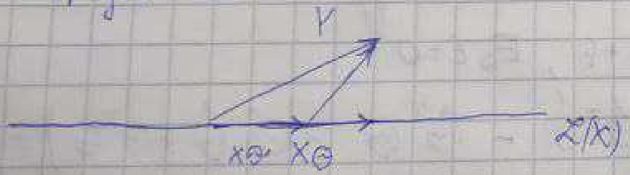
Задача ридж-регресии: $\|Y - X\Theta\|_2^2 + \lambda\|\Theta\|_2^2 \to \min_\Theta$

как было сказано на лекции

$$\begin{cases} \|Y - X\Theta\|_2^2 \to \min_\Theta \\ \|\Theta_1\|_2^2 \le \eta \end{cases}$$

т.е теперь найдя решение задачи регресии нам останется сжать вектор так, чтобы $\|\Theta\|_2^2 \le \eta$. Это и будет решением задачи ридж-регресии

Нарисуем это.



$\|\Theta'\|_2^2 \le \eta$

Также показать требуемое можно зная, что $dist(Y, \mathcal{L}'(X))$ - достигается, т.к

$$\mathcal{L}'(X) = \{X\Theta \mid \Theta \in \mathbb{R}^d, \|\Theta\|_2^2 \le \eta\}$$

$\mathcal{L}'(X)$ - компакт (и не лин подпр-во), а $dist$ - непр.

$\Downarrow$

~~нет~~ $\hat{Y}$ не обязательно $\perp$ $e = Y - \hat{Y}$.