P1.1 For any $\vec{x} = [x_1, \cdots, x_n]^T$, $\vec{y} = [y_1, y_2, \cdots, y_n]^T \in \mathbb{R}^n$ and $\lambda \in (0,1)$

$$\max(\lambda \vec{x} + (1-\lambda)\vec{y}) = \max_{1 \le i \le n} [\lambda x_i + (1-\lambda) y_i]$$

$$\le \max_{1 \le i \le n} \lambda x_i + \max_{1 \le i \le n} (1-\lambda) y_i \overset{\lambda, 1-\lambda > 0}{=} \lambda \max(\vec{x}) + (1-\lambda)\max(\vec{y})$$

Therefore, max is the convex function

The best strategy's $PNL = \max_{1 \le i \le n} \mathbb{E}(PNL(f_i))$

$$\overset{\text{Jensen's Inequality}}{\le} \mathbb{E} \max_{1 \le i \le n} PNL(f_i) = \mathbb{E}\, PNL(f_*)$$

P1.2 Therefore, on average, using $PNL(f_*)$ is an overestimate of the best strategy's PNL. (The inequality is an equality ~~iff~~ $PNL(f_i)$ is constant almost surely, which doesn't hold for the PNL formula)

P2.4.

Since we use SVD decomposition, we know

$\hat{X} = U \Sigma V^T$ gives: $U^T U = I$, $V^T = V^T V = I$, $\Sigma^T \Sigma = \Sigma^2$ is a diagonal matrix.

$\hat{X} = X D^{-1} \iff X = \hat{X} D$, D is a diagonal matrix

Hence:

$X^T X \beta = X^T y \Rightarrow \beta = (X^T X)^{-1} (X^T y)$.

we have:

$$\beta = (X^T X)^{-1} X^T y$$
$$= [(\hat{X} D)^T (\hat{X} D)]^{-1} (\hat{X} D)^T y$$
$$= [D^T V \Sigma \underbrace{U^T U}_{I} \Sigma V^T D]^{-1} D V \Sigma U^T y$$
$$= [D^2 \Sigma^2 \underbrace{V V^T}_{I}]^{-1} D V \Sigma U^T y$$
$$= D^{-1} V \Sigma^{-1} U^T y$$

Then we have:

$$\hat{y} = X \beta = \hat{X} D \beta$$
$$= U \cancel{\Sigma} \cancel{V^T} \cancel{D} \cancel{D^{-1}} \cancel{V} \cancel{\Sigma^{-1}} U^T y$$
$$= U U^T y$$