Target: Representor $\Phi: X \to H$ and predictor $\omega: H \to Y$

$$\min_{\substack{\Phi: X \to H \\ \omega: H \to Y}} \sum_{e \in E_{train}} R^e(\omega \circ \Phi)$$

$$\text{s.t.} \quad \forall e \in E_{train}: \quad \omega \in \arg\min_{\omega^*: H \to Y} R^e(\omega^* \circ \Phi)$$

* Optimize both Φ and ω (hard)

Weaken constraints: Take constraints as penalty

$$L(\Phi, \omega) = \underbrace{\sum_{e \in E_{train}} R^e(\omega \circ \Phi)}_{\text{Empirical Risk}} + \lambda \underbrace{D(\Phi, \omega, e)}_{\text{Invariance}}$$

* $\lambda \in [0, +\infty)$, Balancing Empirical Risk and Invariance
* Unknown form of ω (still hard)

Might as well suppose ω linear

Consider single environment e, we have

$$Y^e = \omega \circ \Phi(X^e)$$

Or in matrix form

$$\Phi(X^e) \cdot \omega = Y^e$$

Least square solution

$$\omega_\Phi^e = [\Phi(X^e)^T \Phi(X^e)]^{-1} \Phi(X^e)^T Y^e$$

Transform

$$\Phi(X^e)^T \Phi(X_e) \omega_\Phi^e - \Phi(X^e)^T Y^e = \vec{0}$$

How far with optimal ω? Just define

$$D(\Phi, \omega, e) = \| \Phi(X^e)^T \Phi(X_e) \omega - \Phi(X^e)^T Y^e \|^2$$

Disaster: consider $(k\Phi, \frac{1}{k}\omega)$, $k \to 0$, we have

$$D(\Phi, \omega, e) = 0$$

And notice

$$\omega \circ \Phi = \underbrace{(\omega \circ \varphi^{-1})}_{\omega'} \circ \underbrace{(\varphi \circ \Phi)}_{\Phi'}$$

We might as well fix $\omega = \omega_0$, Thus

$$\begin{aligned}
L_{\omega_0}(\Phi) &= \sum_{e \in E_{train}} R^e(\omega_0 \circ \Phi) + \lambda D_{\omega_0}(\Phi, e) \\
&= \sum_{e \in E_{train}} R^e(\omega_0 \circ \Phi) + \lambda \| \Phi(X^e)^T \Phi(X_e) \omega - \Phi(X^e)^T Y^e \|^2 \\
&= \sum_{e \in E_{train}} R^e(\omega_0 \circ \Phi) + \lambda \cdot \frac{\partial}{\partial \omega} \left[\tfrac{1}{2} (\Phi(X^e)\omega - Y)^T (\Phi(X^e)\omega - Y) \right]_{\omega_0} \\
&= \sum_{e \in E_{train}} R^e(\omega_0 \circ \Phi) + \lambda \| \nabla_\omega R^e(\omega \circ \Phi) \|^2_{\omega_0}
\end{aligned}$$