Lesson: recap

	Objective	**Solution:** closed-form	**Solution:** via algorithm	**Effect**
LS	$\min_{\beta} \|\|y - X\beta\|\|_2^2$	$\hat{\beta} = (X^T X)^{-1} X^T y$	gradient descent	vanilla
Ridge	$\min_{\beta} \|\|y - X\beta\|\|_2^2 + \lambda \|\|\beta\|\|_2^2$	$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y$	gradient descent	shrinkage
LASSO	$\min_{\beta} \|\|y - X\beta\|\|_2^2 + \lambda \|\|\beta\|\|_1^1$	unavailable	proximal gradient (ISTA): $\beta_{t+1} = ST_{\lambda\eta} \circ \left(\beta_t + \eta \cdot X^T(y - X\beta)\right)$	sparsity
Robust	$\min_{\beta} \rho(y - X\beta)$ e.g. $\rho = \|\|.\|\|_1^1$	unavailable	IRLS: ← cf. cours $D_t = Diag(\omega(y - X\beta_t))$ $\beta_{t+1} = (X^T D_t X)^{-1} X^T D_t y$	robustness