	Models, Assumptions, Goals	Methods	Loss when **learning** using training data	Loss when **predicting** future/test data
I. Probability-Model-based (main, but not the only, focus in traditional statistics) •**Strong** assumptions about probabilistic model, **weak** assumptions about loss	general model: (including **generative** model): $\mathcal{P} = \{p_\theta(X, Y)\}$ **discriminative** model: $\mathcal{P} = \{p_\theta(Y \mid X)\}$ assumption: $\text{'true'} p^* \in \mathcal{P}$ **goal:** use training data to find $\hat{p} \approx p^*$	general learning: ML MAP many others	log-loss penalized log-loss	*Any L((x,y),a) is suitable* if sample size large enough, model assumption correct, and no asymptotic overfitting
		discriminative learning: conditional ML conditional MAP many others *In discriminative models, ML/MAP = conditional ML/MAP*	conditional log-loss cond. pen. log-loss	*Any L(y,a) is suitable* If sample size large enough, model assumption correct and no asymptotic overfitting
II. Function-Model based •**Strong** assumptions about "structural" pattern in data, very weak assumption about noise (iid)	model: $\mathcal{F} = \{f_\theta : \mathcal{X} \to \mathcal{Y}\}$ assumption: $Y = f^*(X) + \text{noise}$ i.e. $E_{p^*}[Y \mid X] = f^*(X)$ with $f^* \in \mathcal{F}$ **goal:** use training data to find $\hat{f} \approx f^*$	least squares regression penalized least squares	mean squared error penalized mean squared error	*Any distance function* $d(f^*, \hat{f})$ *is suitable* if sample size large enough, model assumption correct and no asymptotic overfitting
III. Pure Statistical Learning •**Strong** assumptions about loss of interest (must be known), **weak** assumption on prob.model (only i.i.d.)	model: $\mathcal{F} = \{f_\theta : \mathcal{X} \to \mathcal{Y}\}$ assumption: $\tilde{f} = \arg\min_{f \in \mathcal{F}} \text{EPE}(f) = \arg\min_{f \in \mathcal{F}} E_{P^*}[L^*(Y, f(X))]$ **goal:** use training data to find $\hat{f} \approx \tilde{f}$	Empirical Risk Minimization (ERM) penalized Empirical Risk Minimization (others)	loss L^* of interest or ***proxy*** for L^* penalized L^* or penalized ***proxy*** for L^*	L^*