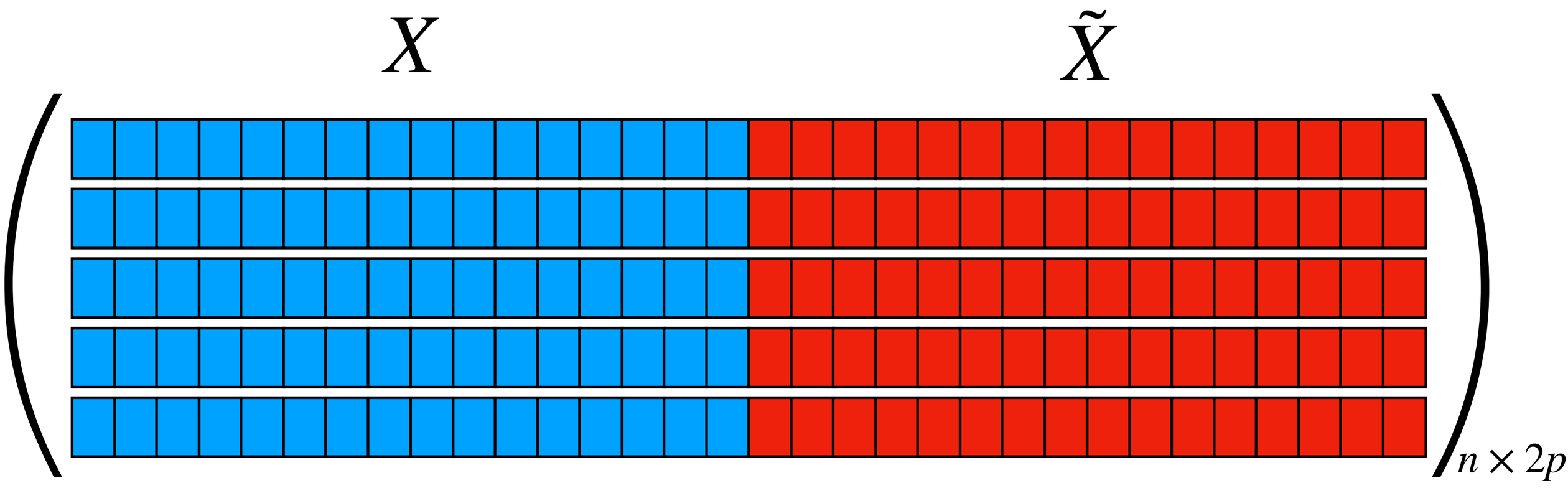
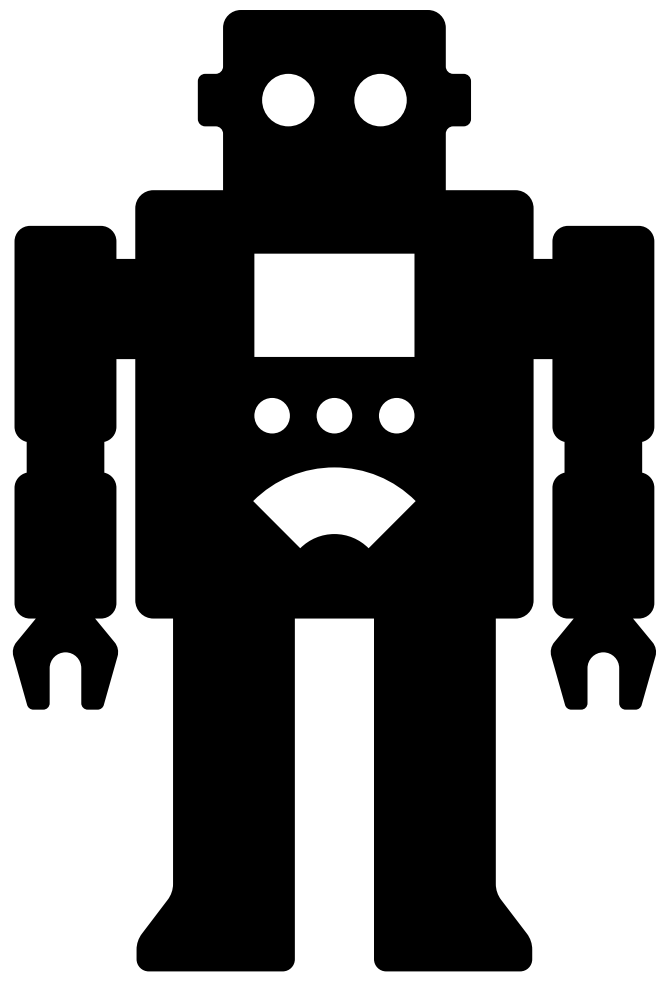
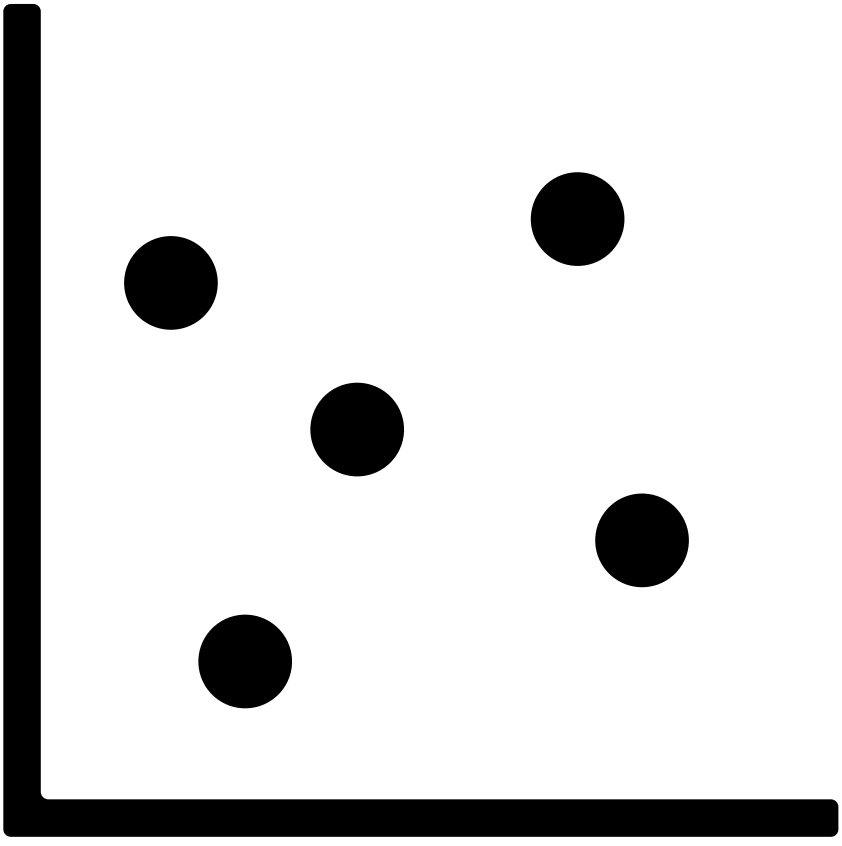
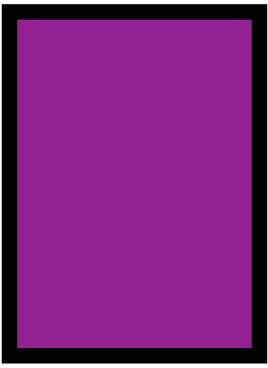
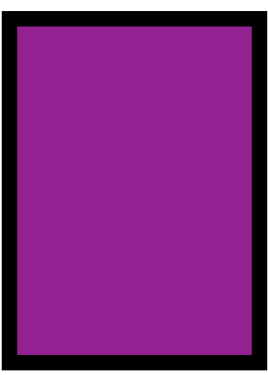
Model-X knockoffs

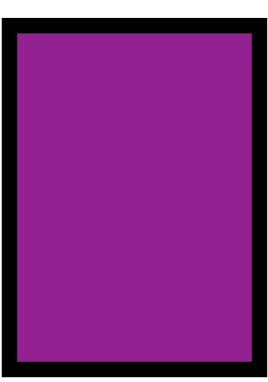

X
$\tilde{X}$
$n \times 2p$

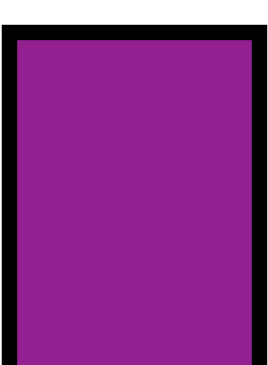


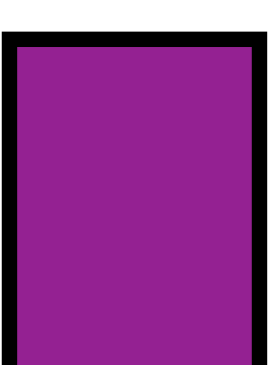


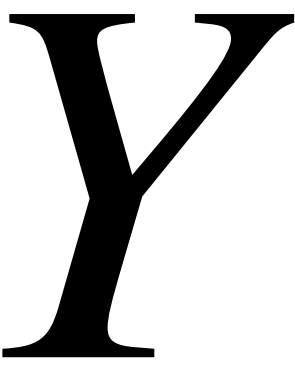


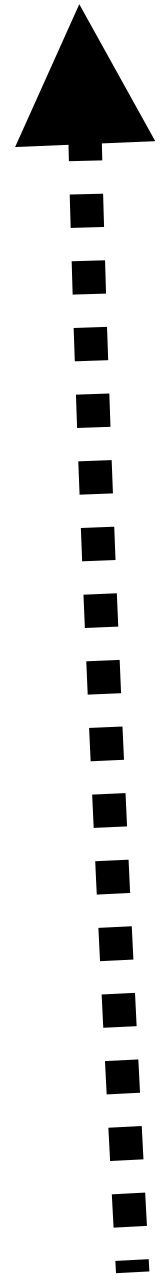




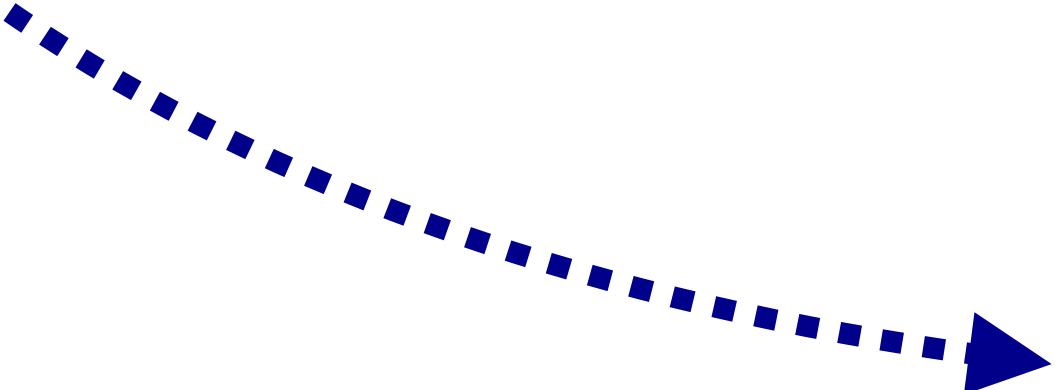








$$T_1, T_2, T_3, T_4, \ldots, T_p, \tilde{T}_1, \tilde{T}_2, \tilde{T}_3, \tilde{T}_4, \ldots, \tilde{T}_p$$



- if X_j is null (not important), can show $T_j = \tilde{T}_j$

- if X_j is non-null (important), want $T_j \gtrapprox \tilde{T}_j$

- any distribution of Y given X

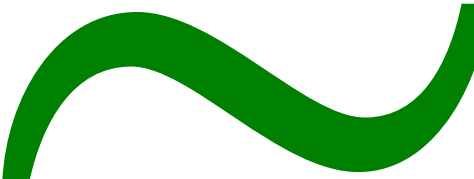
- any feature importance statistics

- finite-sample exact error rate control

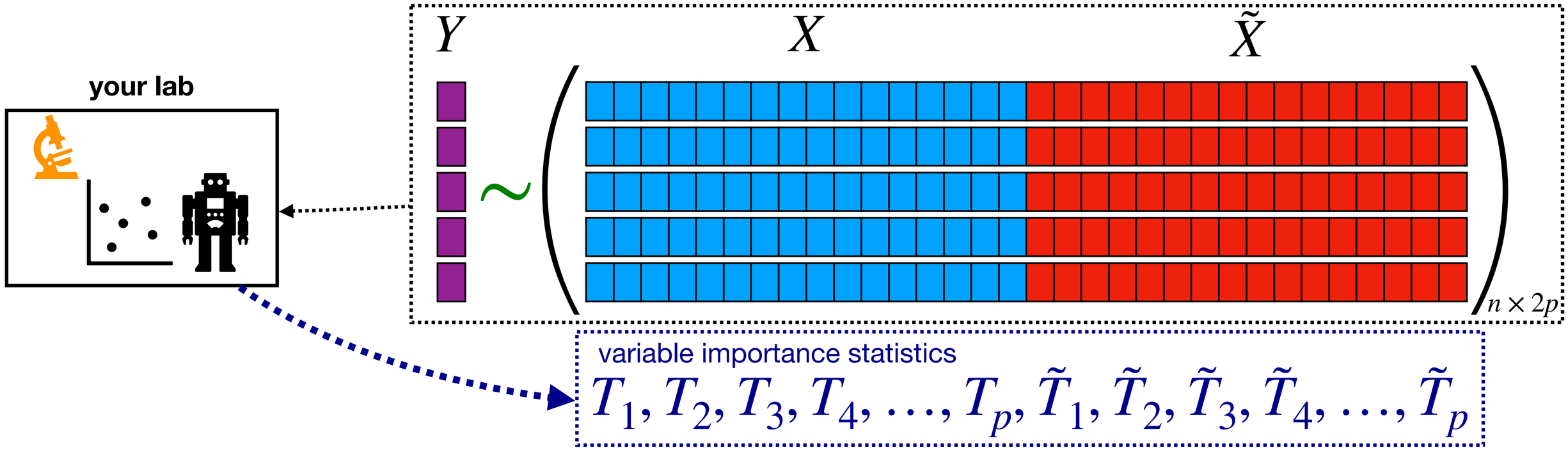
want dependence between $\tilde{X}_j$ and X_j to be low for power

your lab

variable importance statistics



Model-X knockoffs



- any distribution of Y given X
- any feature importance statistics
- finite-sample exact error rate control

want dependence between $\tilde{X}_j$ and X_j to be low for power

- if X_j is null (not important), can show $T_j \overset{d}{=} \tilde{T}_j$
- if X_j is non-null (important), want $T_j \gtrapprox \tilde{T}_j$

Knockoff sampling is difficult

even in the simplest case where distribution of X is known