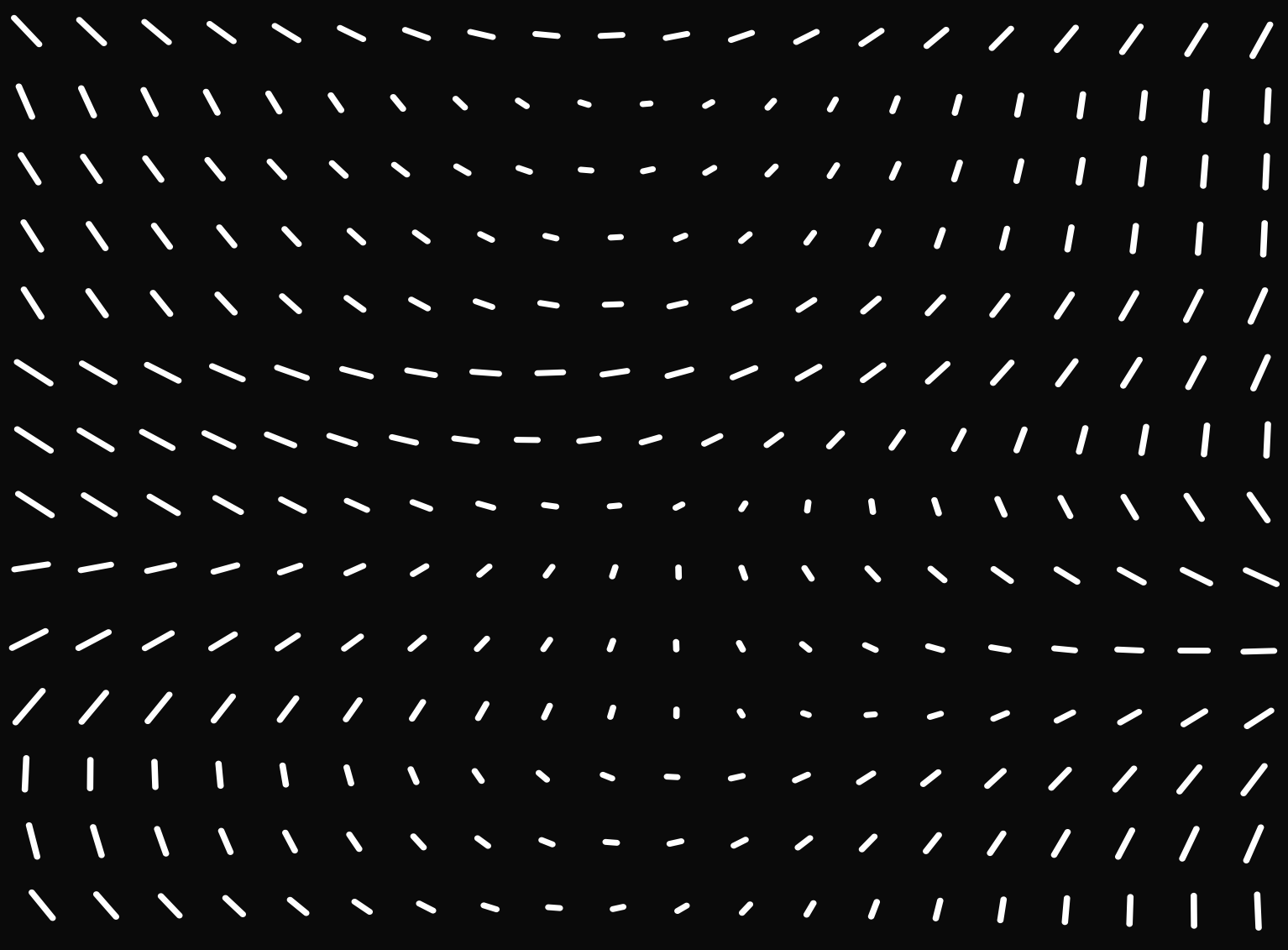
Random Forest & Stuff



1. bootstrap and boosting

2. Random Forest. Modification of bagging

trees:
- can capture complex interaction structure
- if grown sufficiently deep, can have low bias
- noisy, so will benefit great from average
- each tree in RF is identical, so $E(\text{average}) = E(\text{itself})$

2.1 variance reduction

B i.d. trees with correlation ρ and variance σ^2: $\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$

when $B \to \infty$, $\frac{1-\rho}{B}\sigma^2 \to 0$, only $\rho\sigma^2$ left

ρ will affect averaging B trees (bagging)

RF: reduce ρ while control σ^2

2.2. classification/regression and recommendations

input at each split has p variables,
choose $m \le p$ at random as candidates for splitting

- classification: obtain a class vote from each tree, then use majority
  ($m = \sqrt{p}$ and minimum node size is 1)
- regression: average
  ($m = p/3$ and minimum node size is 5)