2.1 rv

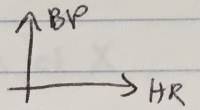
$X \in \mathcal{X} \subseteq \mathbb{R}^p$

$\mathcal{X}$: set of all possible data you could get

$x = [HR, BP]$

$\mathcal{X} = \mathbb{R}^+ \times \mathbb{R}^+$ can't have neg HR/BP

say $p = 2$, $\mathbb{R}^2$



for most datasets p is big ≥100

$y \in \mathcal{y} \subseteq \mathbb{R}^q$

Classification: $\mathcal{y} = \{0, 1, 2\}$

Reg: $\mathcal{y} = \mathbb{R}^2$ ← [age, income]

Dataset

$D_n = \{(x_i, y_i)\}_{i=1}^{n} = \{(x_1, y_1), \dots (x_n, y_n)\}$

(n: # patients)

$$x = \begin{bmatrix} — x_1 — \\ — x_2 — \\ — x_n — \end{bmatrix} \} n$$

(p)

$h(\cdot \mid D_n) : \mathcal{X} \to \mathcal{y} \sim h : x \to y$ (the arguments you choose)

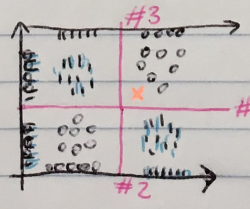
h is hypothesis that takes any possible input and gives any possible output

Integrator: $\mathbb{I}_{x=42}(x) = \begin{cases} 1 & \text{if } x = 42 \\ 0 & \text{if } x \neq 42 \end{cases}$ not important

2.2 Random Forest

$x \in \mathcal{X} \subseteq \mathbb{R}^2$

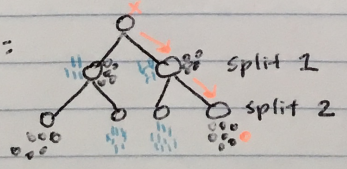
$y \in \mathcal{y} \subseteq \{0, 1\}$



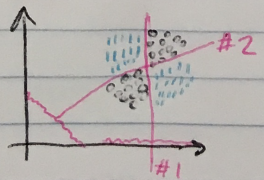
split after projecting data onto axis

*can't project on axis again after doing it once

tree version:



SPORF allows projection on diagonal lines



RF can only split ↑ ↓ → ←

SPORF Paper Notes

2.1 Statistical Learning: Classification

$X \in \mathbb{R}^p$

$Y \in \mathcal{Y} = \{C_1, \ldots C_K\}$ — class labels of X

(X, Y) has joint distribution f_{XY}

$(X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)$ sampled iid

Arbitrary training set: $\mathcal{D}_n = \{X_i, Y_i\}_1^n$

Learn classifier $h(\cdot\,; \mathcal{D}_n) : \mathbb{R}^p \to \mathcal{Y}$ predicts Y (unobserved) given X (observed)

↳ $\min P_{XY}(h(X; \mathcal{D}_n) \neq Y)$ probability of misclassification

ex) Bayes classifier: $h^*(X) = \underbrace{\operatorname{argmax}}_{\text{generally unknown in real-world}} \underbrace{P_{Y|X}}_{\text{generally unknown in real-world}} (Y = k \mid X)$

1. Axis aligned vs Oblique

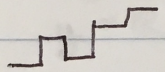
Axis aligned: feature space recursively split along directions parallel to coordinate axis

Oblique: better learning properties and expressive capacity

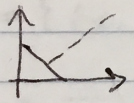
↳ SPORF searches splits over sample of very sparse random projections

2.3 Oblique Extensions to Random Forest

- RF limited since splits must be along coord. axis



- Oblique decision forests can split along directions oblique to axes.



- ex) Forest-RC
  - d univariate projections, each a linear combo of L randomly chosen dimensions.
  - Weight of each proj. independently sampled uniformly over interval $[-1, 1]$

2.4 Random Projections

Construct random projection matrix $A \in \mathbb{R}^{p \times d}$, multiply by X

$\tilde{X} = XA \in \mathbb{R}^{n \times d}$, $d \ll \min(n, p)$

↳ Preserves pairwise distances of X mostly

Very sparse random Projections

$$a_{ij} = \begin{cases} +1 & \text{prob } \frac{1}{2s} \\ 0 & \text{prob } 1 - \frac{1}{s} \\ -1 & \text{prob } \frac{1}{2s} \end{cases} \qquad \text{typically } s \gg S$$