Lecture 9/10: Prediction and Overfitting

Tuesday, September 10, 2019 2:43 PM

Deterministic label: The target (f) is deterministic ← previous

---

What is the prob. of seeing a tasty apple and has feature x_1

$$P(X=x_1, Y=yes) = P(X=x_1) \cdot P_r(Y=yes \mid X=x_1)$$

$$0.25 \cdot 0.9 =$$

$x_3 = (B, red, medium, Crunchy)$ $\quad P(X=x_3) = 0.3$

Prob of seeing apple x_3 that was also tasty $= 0.2$

$$P_r(Y=yes \mid X=x_3) = \frac{P_r(X=x_3, Y=yes)}{P_r(X=x_3)} = \frac{0.2}{0.3} = 2/3$$

$$P_r\left(S=\{(x_1,y_1)\}\right) = P(X=x_1, Y=y_1)$$

$$P_r\left(S=\{(x_1,y_1),(x_2,y_2)\}\right) = P(X_1=x_1, Y_1=y_1) \cdot P\left(X_2=x_2, Y_2=y_2 \mid X_1=x_1, Y_1=y_1\right)$$

(independence) $= (X_1=x_1, Y_1=y_1) \cdot P(X_2=x_2, Y_2=y_2)$

(Identically)

$$= P(X=x_1, Y_1=y_1) \cdot P(X=x_2, Y=y_2)$$

$$Pr\left(S=\{(x_1,y_1)\cdots(x_m,y_m)\}\right) = \prod_{i=1}^{m} P(X=x_i, Y=y_i) \cdot$$

$$\mathop{\mathbb{E}}_{(x,y)\sim P}\left[\Delta(h(x),y)\right] = \mathop{Pr}_{(x,y)\sim P}\left[h(x)\neq y\right] = \sum_{\substack{x\in X \\ y\in Y}} \Delta(h(x),y)\cdot P(X=x, Y=y)$$

$err_{S_{train}}(h_{S_{train}}) = 0$

$err_{P}(h_{S_{train}}) = 1/2$ large.

Overfitting:

* S too small.
* X is rich, H is rich
* Memorize S without "learning" patterns.

DTs are expressive

$H: 2^X$ H: AND OR Decision lists ...

Assumption:

* Distances.
* Features meaningful
* Assumption on H.

$|H| = 2^{|X|}$ 2^{24} all function

Occam's Razor:

Good: World is simple, structured.

Simple → faked.

Bad World is Not simple. ←

Simplicity perpetuating stereotypes / biases.