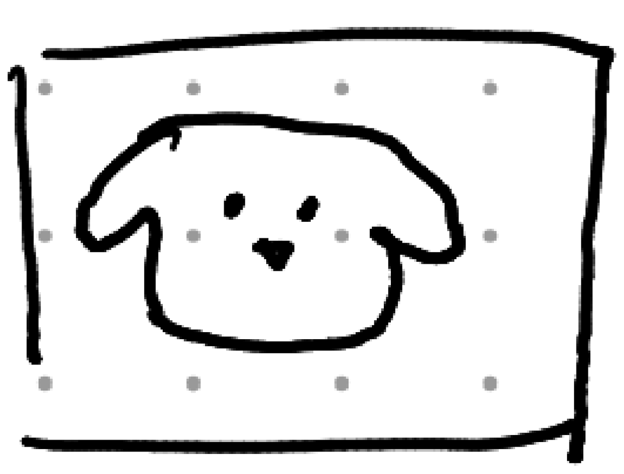
Regression / Classification

Data comes in the form

$$(X_1, Y_1), \ldots, (X_n, Y_n)$$

Ex.

$X_1 =$  $Y_1 =$ "cat"

$X_2 =$  $Y_2 =$ "dog"

$\vdots$ $\vdots$

Ex. $X_i =$ age, GPA, exam scores, gender, class list, etc.

$Y_i \in$ {"interview", "reject"}

Ex. $X_i =$ { home team record, salaries, etc. / away team record, salary, etc. }

$Y_i =$ point margin

We will view (X, Y) as an iid sample from some distribution $F_{X,Y}$

We would like $Y = r(X)$, but this is not possible, because same input might result in different outputs.

Instead, we might try to compute sth like

$$r(x) = \mathbb{E}[Y \mid X = x] \quad \text{(regression fxn)}$$

Note

Even if we had a perfect algorithm to compute $r(x)$, our regression fxn will reflect biases in the data

Ex. - Biases in hiring (names)

- Biases in allocating resources

- Etc.

Solving the math problem $\neq$ solving real world problem..

How do we find an approximate regression fxn?

Let $\hat{r}_\beta(x)$ be a function depend on parameters β

Ex. $\hat{r}_\beta(x) = \beta_0 + \beta_1 x$ $\qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$

$\hat{r}_\beta(x) = \sigma(Wx + b)$ $\qquad \beta = (W, b)$

we want to find parameters β which minimize

$$R(\beta) = \mathbb{E}\left[L(\hat{r}_\beta(X), Y)\right]$$

where $L(\hat{y}, y)$ is a "loss function"

Ex. $$L(\hat{r}_\beta(X), Y) = \begin{cases} 1 & \hat{r}_\beta(X) \neq Y \\ 0 & \hat{r}_\beta(X) = Y \end{cases}$$

$$L(\hat{r}_\beta(X), Y) = |\hat{r}_\beta(X) - Y|^2$$

We cant compute this expectation bc. we don't know $F_{X,Y}$

But, Given data $(X_1, Y_1), \ldots, (X_n, Y_n)$ we can compute empirical risk

$$R_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} L(\hat{r}_\beta(X_i), Y_i)$$

<u>Linear regression</u>

$\hat{r}_\beta(x) = \beta_0 + \beta_1 x \qquad L(\hat{y}, y) = (\hat{y} - y)^2$

Minimize $\quad R_n(\beta) = \sum_{i=1}^{\hat{}} (\beta_0 + \beta_1 X_i - Y_i)^2$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}$$

$$\hat{\beta}_0 = \overline{Y}_n + \hat{\beta}_1 \overline{X}_n$$