$p(\underbrace{\text{error}}_{\hat{y} = \hat{f}(x) \neq y^*} \mid \text{type} = A, t)$ ← some fixed threshold; $\hat{y} = \mathbb{1}[x \geq t]$ — threshold classifier error

expected error: average error over all possible test data points

$$\mathbb{E}_{x, y^* \sim p^*(x,y)} \left[p(\hat{f}(x) \neq y^*) \mid \text{type} = A, t \right]$$

$$\sum_{k=0}^{1} \int p^*(x, y=k)\, p(\hat{f}(x) \neq k \mid \text{type} = A, t)\, dx = \left(t - \frac{1}{2}\right)^2 + \frac{1}{4}$$

classes are discrete $\Rightarrow \mathbb{E} \Rightarrow \sum$

features are continuous $\mathbb{E} \Rightarrow \int$

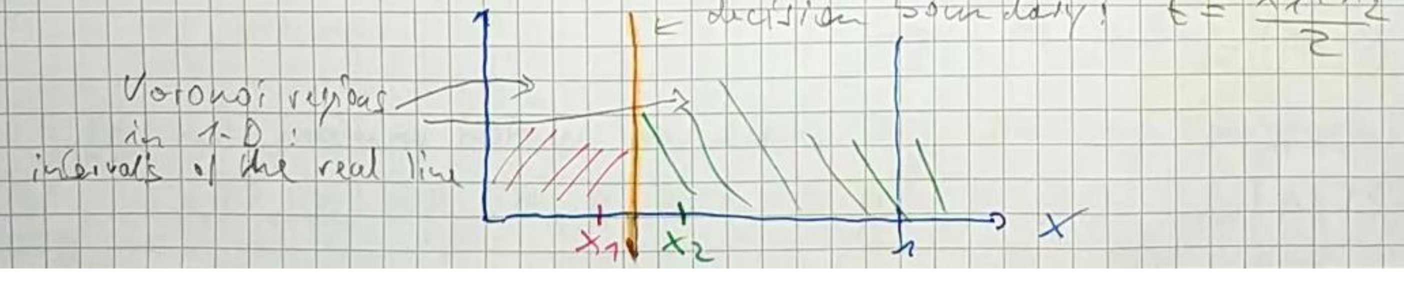
---

apply this result to the nearest-neighbor classifier for $N = 2$

- we do not fix t manually, but by means of the TS: $t = \frac{x_1 + x_2}{2}$

$\{(x_1, y_1 = 0), (x_2, y_2 = 1)\}$

if $x_1 < x_2 \Rightarrow$ we get a type A threshold classifier, $x_1 > x_2 \Rightarrow$ type B



to compute the expected error of the nearest neighbor classifier for $N = 2$, we must take two averages:
- average about all possible test points (as before)
- average over all possible training set, because the classifier performance depends on the training set

e.g. $TS_1 = \{ (x_1 = 0.2, y_1 = 0), (x_2 = 0.8, y_2 = 1) \}$

good $\quad t = \frac{0.2 + 0.8}{2} = 0.5, \quad x_1 < x_2 \Rightarrow$ type A

$TS_2 = \{ (x_1 = 0.5, y_1 = 0), (x_2 = 0.4, y_2 = 1) \}$

bad $\quad t = \frac{0.5 + 0.4}{2} = 0.45, \quad x_1 > x_2 \Rightarrow$ type B

$$\mathbb{E}_{TS} [\underbrace{p(\text{error} \mid TS)}_{= p(\text{error} \mid \text{type}, t)}]$$ with type and t calculated from TS

in general $\int p(TS) \cdot p(\text{error} \mid TS) \, dTS$

i.i.d. assumption: independent, identically distributed

TS has two properties: (1) each training instance was obtained independently of the other training instances

dice: i-th throw outcome independent of $i' < i$ outcomes

$$\Rightarrow p(TS) = p(\{(x_i, y_i)\}_{i=1}^{N}) \text{ factorizes}$$
$$= p_1(x_1, y_1) \cdot p_2(x_2, y_2) \cdot \ldots \cdot p_N(x_N, y_N)$$

(2) each training instance was drawn from the same (unknown) true probability distr.

$$p_1(x_1, y_1) = p_2(x_2, y_2) = \ldots = p^*(x, y) \Rightarrow \boxed{p(TS) = \prod_{i=1}^{N} p^*(x_i, y_i)}$$

for all i: use the same dice

$$\mathbb{E}_{TS}\left[p(\text{error} \mid TS)\right] = \int p(TS)\, p(\text{error} \mid TS) \, dTS$$

$$\overset{\text{i.i.d.}}{=} \int p(x_1, y_1) \cdot p(x_2, y_2) \cdots p(x_N, y_N)\, p(\text{error} \mid \{(x_i, y_i)\}_{i=1}^{N})$$

$$d(x_1, y_1) \cdot \ldots d(x_N, y_N)$$

specifically for the toy example : we already know that $y_1 = 0$, $y_2 = 1$

$\Rightarrow$ the integration and prob of y can be dropped

$$p(TS) = p(x_1 \mid y_1 = 0) \cdot p(x_2 \mid y_2 = 1)$$

i.i.d assumption applied to TS $\{(x_1, y_1 = 0), (x_2, y_2 = 1)\}$

if in addition, $x_1 < x_2$ $\Rightarrow$ we get a type A classifier

$$\mathbb{E}_{\substack{TS \\ x_1 < x_2}}\left[p(\text{error} \mid TS)\right] = \int_0^1\int_0^1 p(x_1 \mid y_1 = 0) \cdot p(x_2 \mid y_2 = 1)\, \mathbb{1}[x_1 < x_2]\, p\left(\text{error} \mid \text{type } A, t = \frac{x_1 + x_2}{2}\right) dx_2\, dx_1$$

$$= \int_0^1 \underbrace{p(x_1 \mid y_1 = 0)}_{2 - 2x} \left[\int_{x_1}^1 \underbrace{p(x_2 \mid y_2 = 1)}_{2x}\, \underbrace{p\left(\text{error} \mid \text{type} = A, t = \frac{x_1 + x_2}{2}\right)}_{(t - \frac{1}{2})^2 + \frac{1}{4} = \left(\frac{x_1 + x_2}{2} - \frac{1}{2}\right)^2 + \frac{1}{4}} dx_2\right] dx_1$$

$\Rightarrow$ give all this to Wolfram Cloud and get $83/360$

$$\mathbb{E}_{\substack{TS \\ x_2 < x_1}}\left[p(\text{error} \mid TS)\right] = 43/360 \qquad p(TS \mid y_1 \neq y_2)$$

$$\mathbb{E}_{TS}\left[p(\text{error} \mid TS)\right] = \mathbb{E}_{\substack{TS \\ x_1 < x_2}} + \mathbb{E}_{x_1 > x_2} = 35\%$$