

CONFORMAL PREDICTION

IA Frameworks

Joseba Dalmau

UNCERTAINTY QUANTIFICATION IN DEEP LEARNING

- Can we trust the predictions of our models ?
- Under what circumstances ?
- To what extent ?

CLASSICAL UQ TECHNIQUES

- Bayesian methods
- Ensembling
- Dropout
- ...



NON POST-HOC

WITHOUT GUARANTEES

CONFORMAL PREDICTION:

OBJECTIVE AND GUARANTEE

Given: a predictor $\hat{f}: \mathcal{X} \rightarrow \mathcal{Y}$
and a nominal error rate α .

Build: $\hat{C}_\alpha: \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$

With the following guarantee:

$$\mathbb{P}(y_{\text{test}} \in \hat{C}_\alpha(x_{\text{test}})) \geq 1 - \alpha$$

OUTLINE

→ Conformal Regression

- Theorem
- Jackknife+ / CV+
- CQR

→ Conformal Classification

- LAC
- APS / RAPS

CONFORMAL REGRESSION

Given: $\hat{f}: X \rightarrow Y$ and α .

Build: $\hat{C}_\alpha: X \rightarrow \mathcal{P}(Y)$

With the following guarantee:

$$\mathbb{P}(Y_{\text{test}} \in \hat{C}_\alpha(X_{\text{test}})) \geq 1 - \alpha$$

CONFORMAL REGRESSION

Given: $\hat{f}: \mathcal{X} \rightarrow \mathbb{R}$ and α .

Build: $\hat{C}_\alpha: \mathcal{X} \rightarrow \underbrace{\mathcal{P}(\mathbb{R})}_{\text{usually an interval!}}$

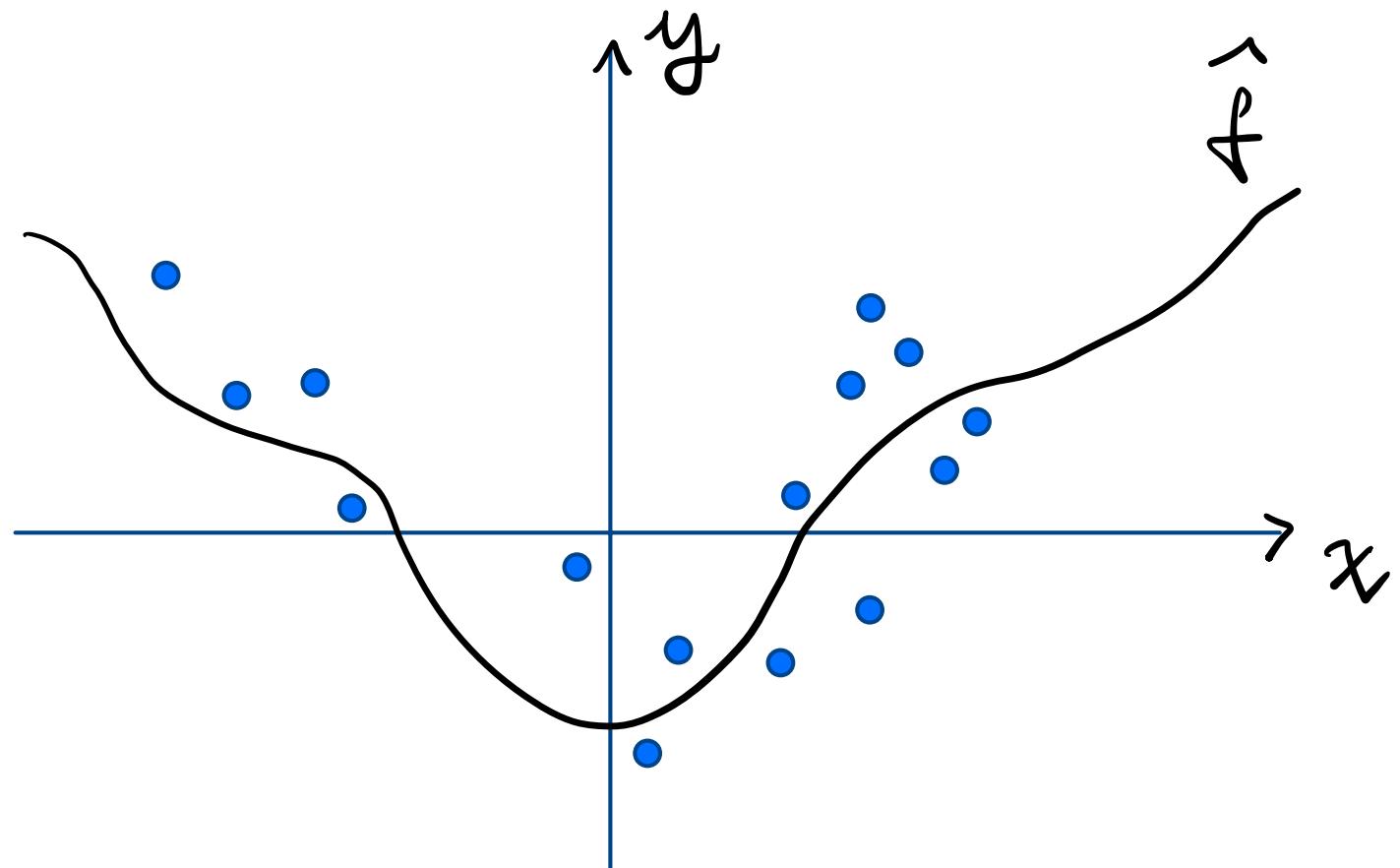
With the following guarantee:

$$\mathbb{P}(Y_{\text{test}} \in \hat{C}_\alpha(X_{\text{test}})) \geq 1 - \alpha$$

CALIBRATION

We use a calibration set

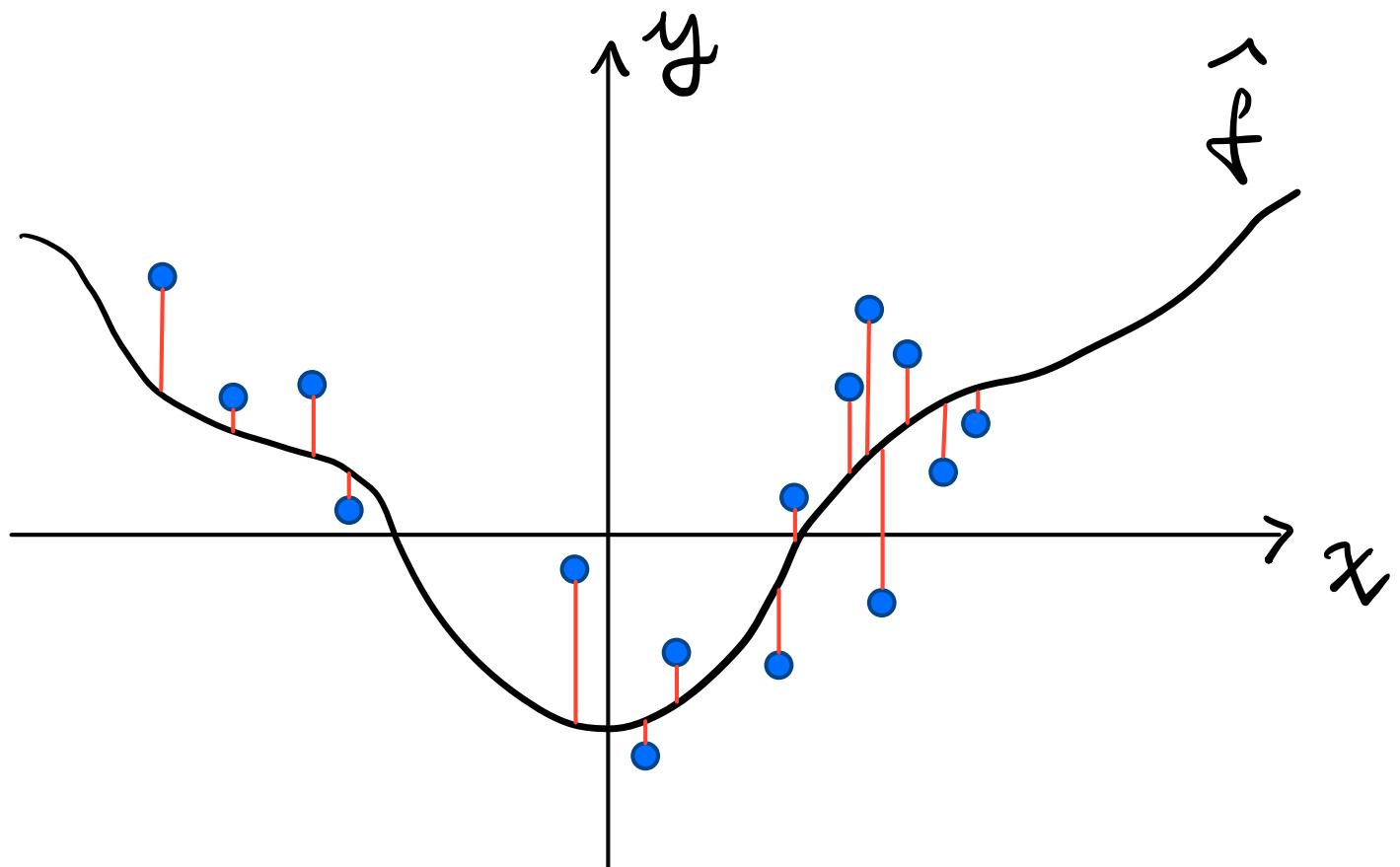
$$D_{\text{calib}} = \{(x_i, y_i)\}_{i=1}^n$$



CALIBRATION

We measure the **scores** (errors)

$$S_i = |y_i - \hat{f}(x_i)|$$



CALIBRATION

We compute:

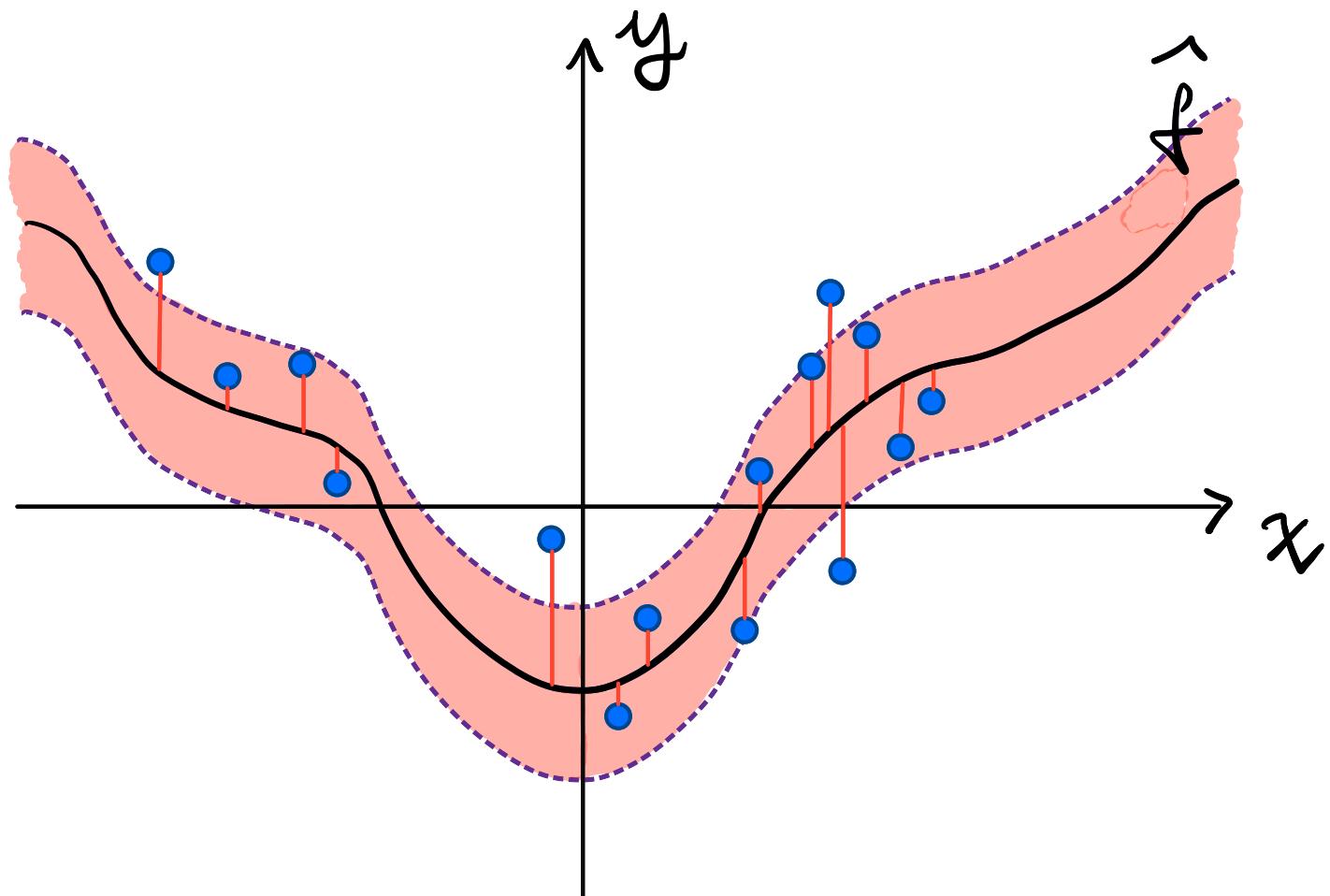
$s_\alpha :=$ the $\frac{\lceil (n+1)(1-\alpha) \rceil}{n}$ -th

quantile of the scores s_1, \dots, s_n

i.e. the $\lceil (n+1)(1-\alpha) \rceil$ -th
smallest score.

CALIBRATION

We predict $\hat{C}_\alpha(x) = [\hat{f}(x) - \delta_\alpha, \hat{f}(x) + \delta_\alpha]$



GUARANTEE

Theorem .- Given a calibration set

$\{(x_i, y_i)\}_{i=1}^n$ and a test point (x_{n+1}, y_{n+1}) ,

if the scores $s_i = |y_i - \hat{f}(x_i)|$ are exchangeable, then the interval

$\hat{C}_\alpha(x_{n+1}) = [\hat{f}(x_{n+1}) - s_\alpha, \hat{f}(x_{n+1}) + s_\alpha]$
satisfies:

$$P(y_{n+1} \in \hat{C}_\alpha(x_{n+1})) \geq 1 - \alpha$$

EXCHANGABILITY

The random variables (X_1, X_2, \dots, X_n) are exchangeable if:

For any $\pi \in \{ \text{permutations of } \{1, \dots, n\} \}$

$$X_{\pi(1)}, \dots, X_{\pi(n)} \stackrel{\text{law}}{=} X_1, \dots, X_n$$

Exercise.— Give an example of an array of r.v.s that are exchangeable but not i.i.d.

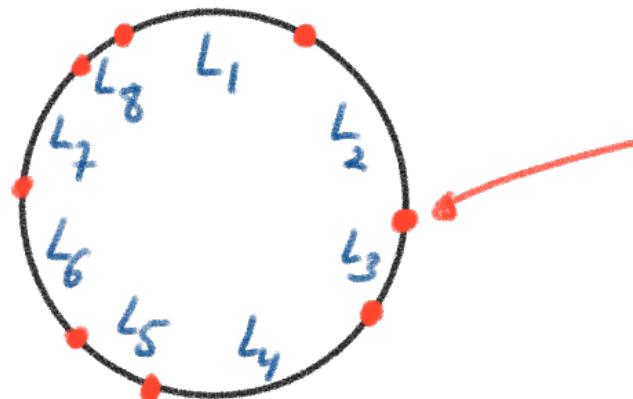
EXCHANGABILITY

The random variables (X_1, X_2, \dots, X_n) are exchangeable if:

For any $\pi \in \{ \text{permutations of } \{1, \dots, n\} \}$

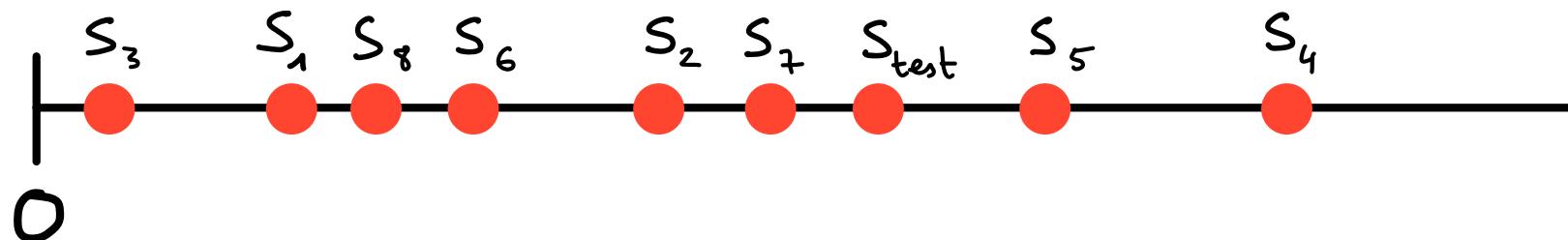
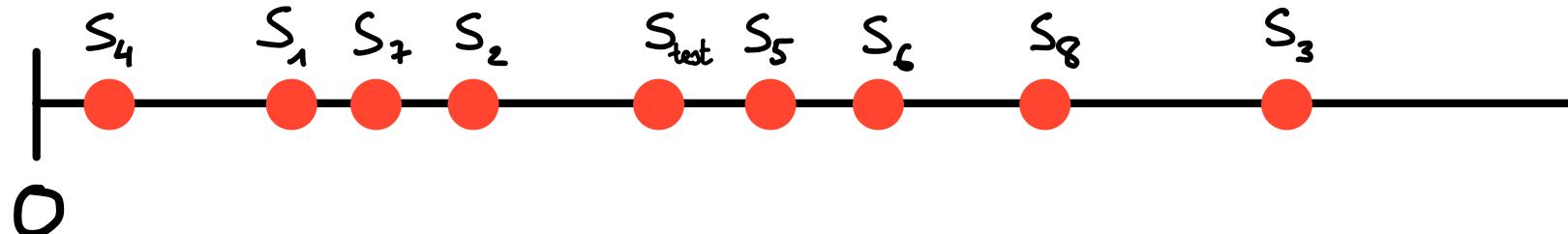
$$X_{\pi(1)}, \dots, X_{\pi(n)} \stackrel{\text{law}}{=} X_1, \dots, X_n$$

Example:

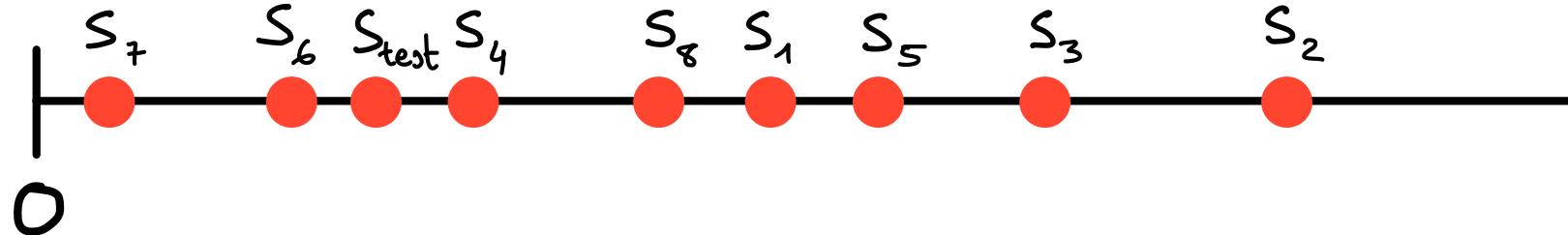


points drawn as
i.i.d. uniform
r.v.s on the S^1

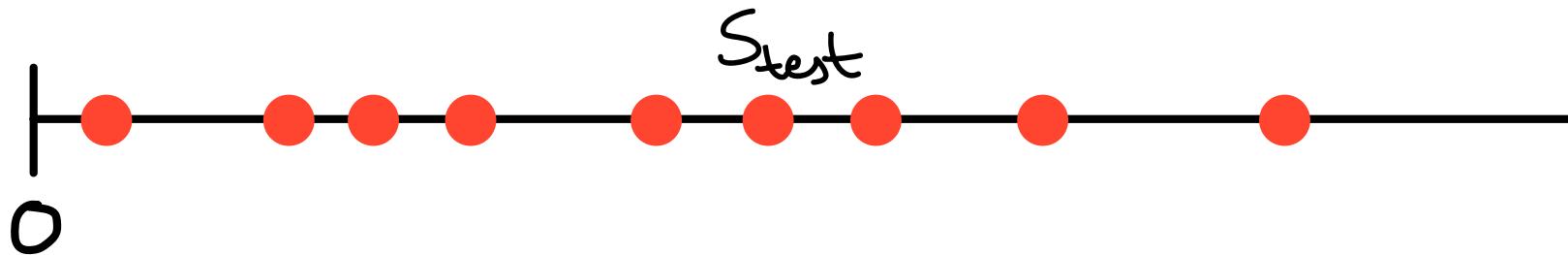
PROOF



⋮



PROOF



$P(\text{Rank of } S_{\text{test}} = k)$

$$= \frac{n!}{(n+1)!} = \frac{1}{n+1}$$

$\Rightarrow \text{Rank of } S_{\text{test}} \sim \text{Unif}(\{1, \dots, n+1\})$

PROOF

Rank of S_{test} $\sim \text{Unif}(\{1, \dots, n+1\})$

$$\Rightarrow P(\text{Rank of } S_{\text{test}} \leq K) = \frac{K}{n+1}$$

We choose the smallest K s.t.

$$\frac{K}{n+1} \geq 1 - \alpha \text{ i.e. } K = \lceil (n+1)(1-\alpha) \rceil$$

Too Good To BE TRUE ?



\hat{f} bad

$\Rightarrow \hat{C}_\alpha$ very large

ADVANTAGES

- Post-hoc
- Distribution-free
- Minimal hypotheses
- Finite-sample guarantee

LIMITATIONS

Calibration- marginal guarantee:

$$\mathbb{P}(Y_{n+1} \in \underbrace{\hat{C}_\alpha(X_{n+1})}_{\text{}}) \geq 1 - \alpha$$

how is this set
constructed?

LIMITATIONS

Calibration - marginal guarantee:

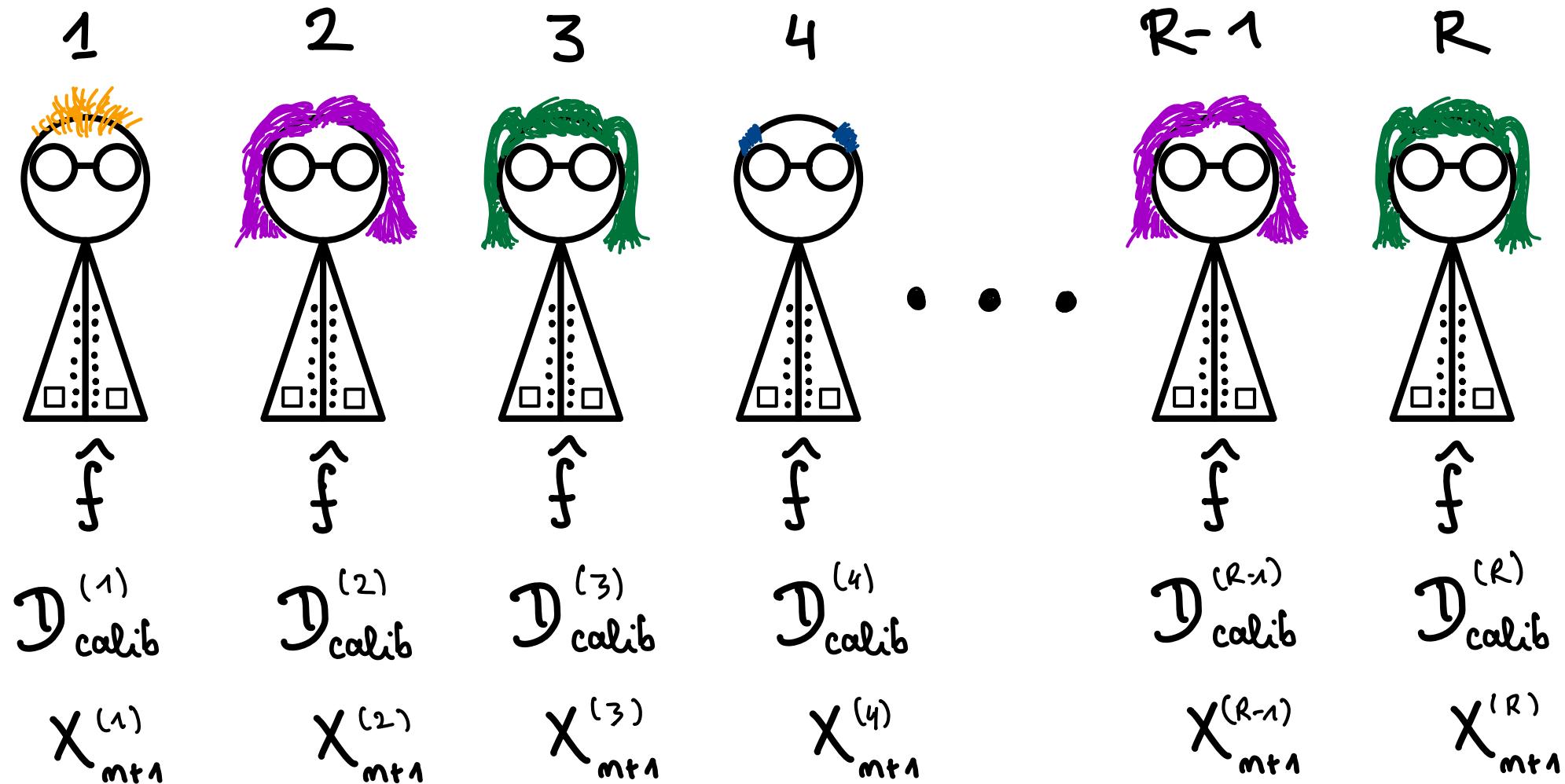
$$\mathbb{P}(Y_{n+1} \in \underbrace{\hat{C}_\alpha(X_{n+1})}_{\text{}}) \geq 1-\alpha$$

how is this set
constructed?

$$\mathbb{P}(Y_{n+1} \in \hat{C}(X_{n+1}) \mid X_1, \dots, X_n) \geq 1-\alpha ?$$

LIMITATIONS

Calibration- marginal guarantee:

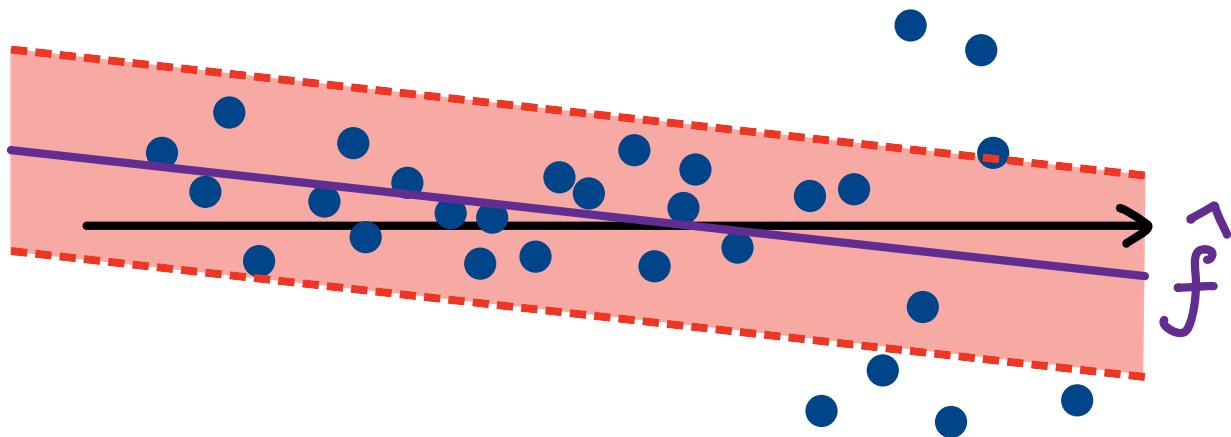


LIMITATIONS

Non-conditional guarantee

$$\mathbb{P}(Y_{n+1} \in \hat{C}_\alpha(X_{n+1})) \geq 1-\alpha \quad \checkmark$$

$$\mathbb{P}(Y_{n+1} \in \hat{C}_\alpha(X_{n+1}) | X_{n+1} = x) \geq 1-\alpha \quad \times$$



SPLIT CONFORMAL

Theorem. - Given a calibration set

$\{(x_i, y_i)\}_{i=1}^n$ and a test point (x_{n+1}, y_{n+1}) ,

if the scores $s_i = |y_i - \hat{f}(x_i)|$ are

exchangeable, then the interval

$$\hat{C}_\alpha(x_{n+1}) = [\hat{f}(x_{n+1}) - s_\alpha, \hat{f}(x_{n+1}) + s_\alpha]$$

satisfies:

$$\mathbb{P}(y_{n+1} \in \hat{C}_\alpha(x_{n+1})) \geq 1 - \alpha$$

SPLIT CONFORMAL

\hat{f} model (obtained by optimizing
on D_{train})

$D^{\text{calib}} = \{ (x_i, y_i) \}_{i=1}^n$ calibration
data

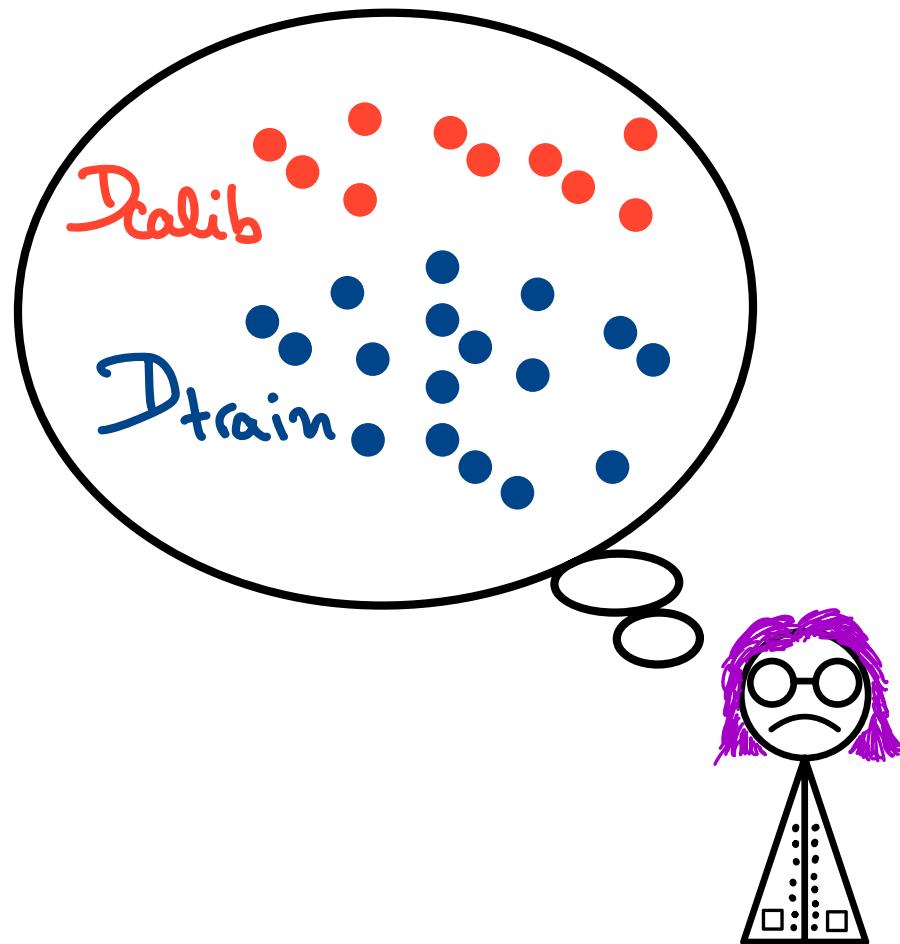
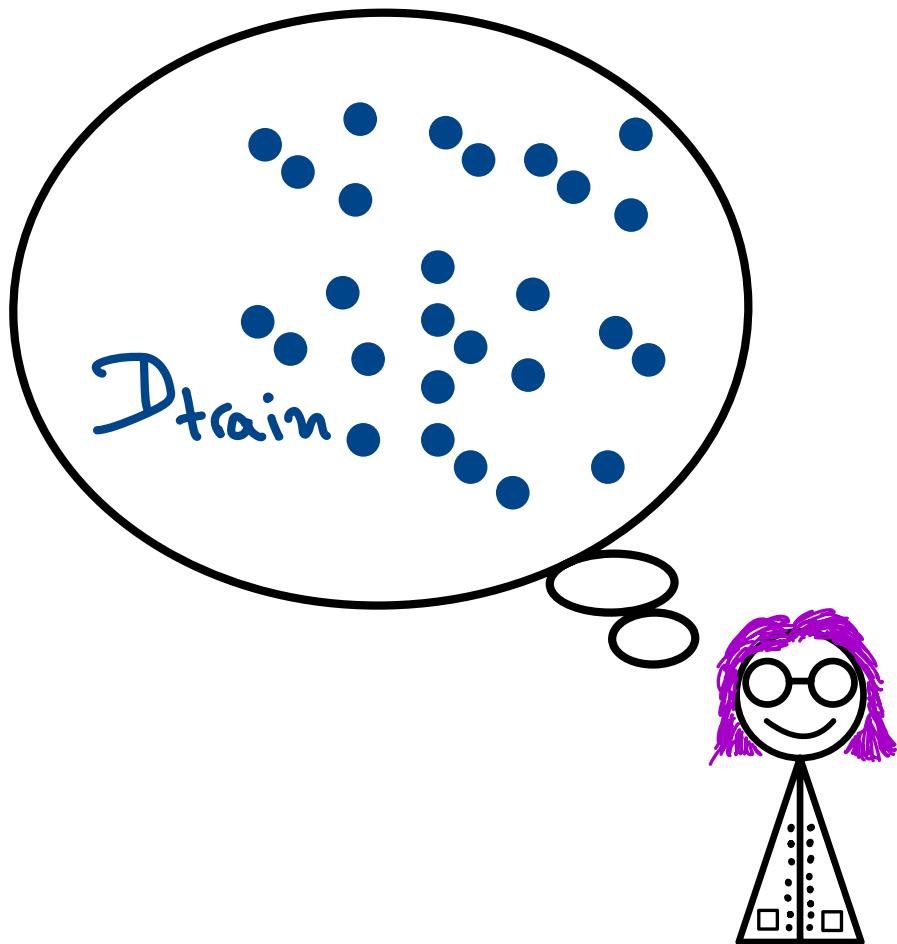
(x_{n+1}, y_{n+1}) test point.



$s_i = |y_i - \hat{f}(x_i)|$, $i=1, \dots, n+1$
exchangeable

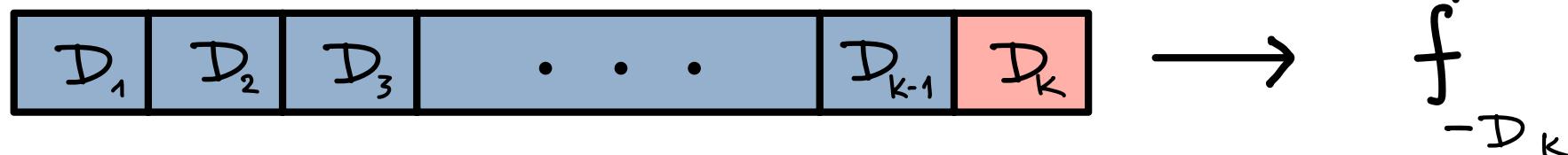
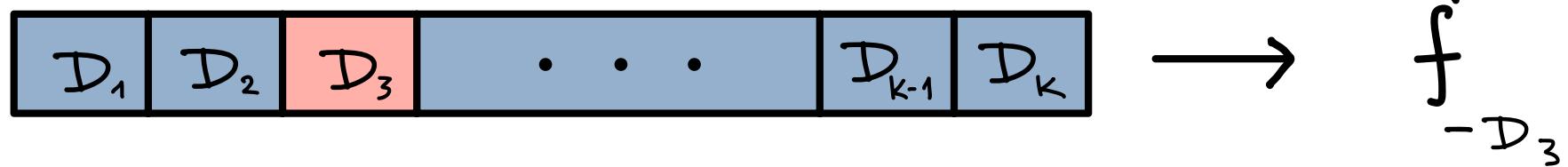
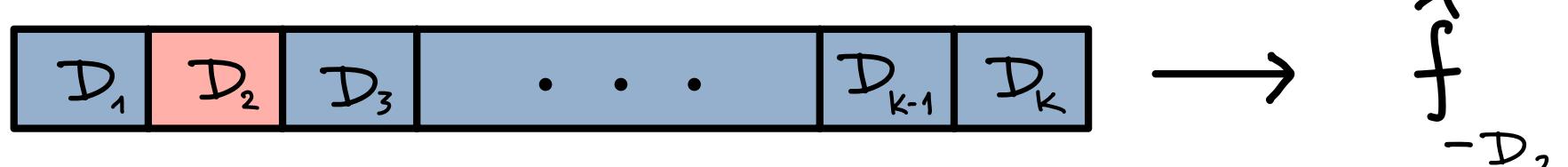
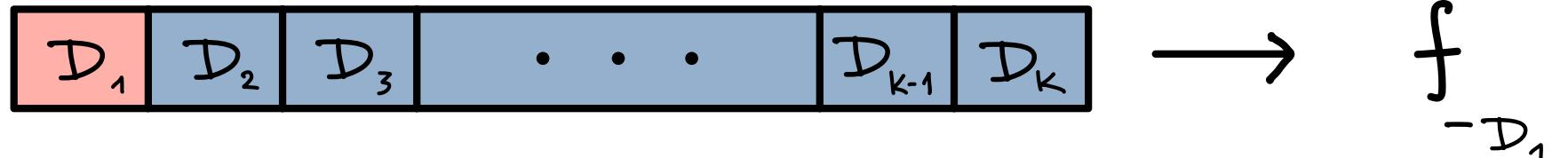
LIMITATIONS

Need for calibration data



JACKNIFE+ , CROSS-VALIDATION+

Partition the data :



JACKNIFE+ , CROSS-VALIDATION+

Calibration:

$$S_i^{cv} = |Y_i - \hat{f}_{-D_{ind(i)}}(X_i)|, \quad i=1, \dots, n$$

JACKNIFE+ , CROSS-VALIDATION+

Inference:

$\hat{e}_\alpha(x) := \lfloor \alpha(n+1) \rfloor$ - the smallest value of

$$\hat{f}_{-S_{\text{ind}(1)}}^{(x)-S_1^{\text{cv}}} \dots \hat{f}_{-S_{\text{ind}(n)}}^{(x)-S_n^{\text{cv}}}$$

$\hat{\mu}_\alpha(x) := \lceil (1-\alpha)(n+1) \rceil$ -th smallest value of

$$\hat{f}_{-S_{\text{ind}(1)}}^{(x)+S_1^{\text{cv}}} \dots \hat{f}_{-S_{\text{ind}(n)}}^{(x)+S_n^{\text{cv}}}$$

$$\hat{C}_\alpha(x) = [\hat{e}_\alpha(x), \hat{\mu}_\alpha(x)]$$

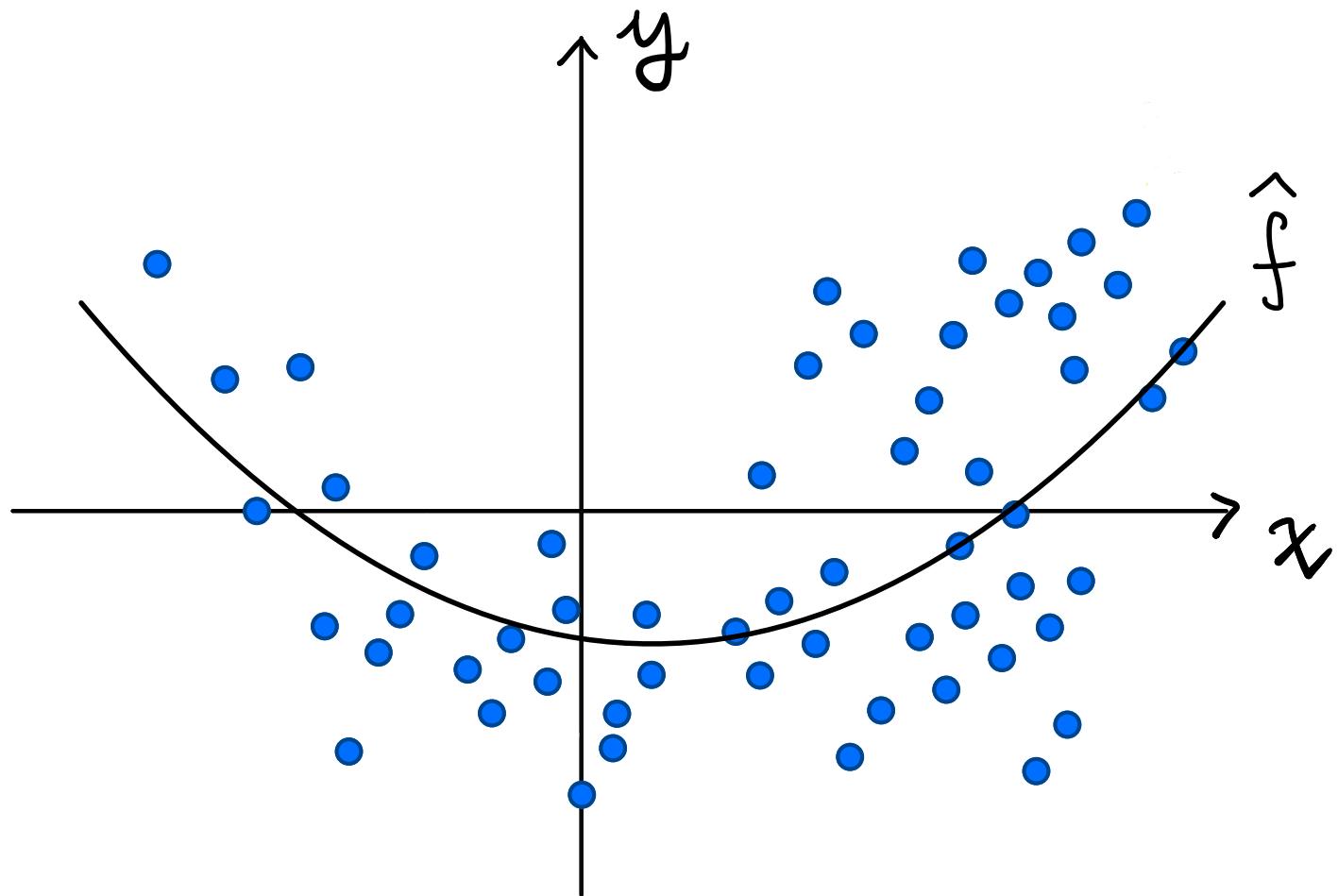
JACKNIFE+, CROSS-VALIDATION+

Guarantee:

$$P(Y_{n+1} \in \hat{C}_\alpha(X_{n+1})) \geq 1 - 2\alpha$$

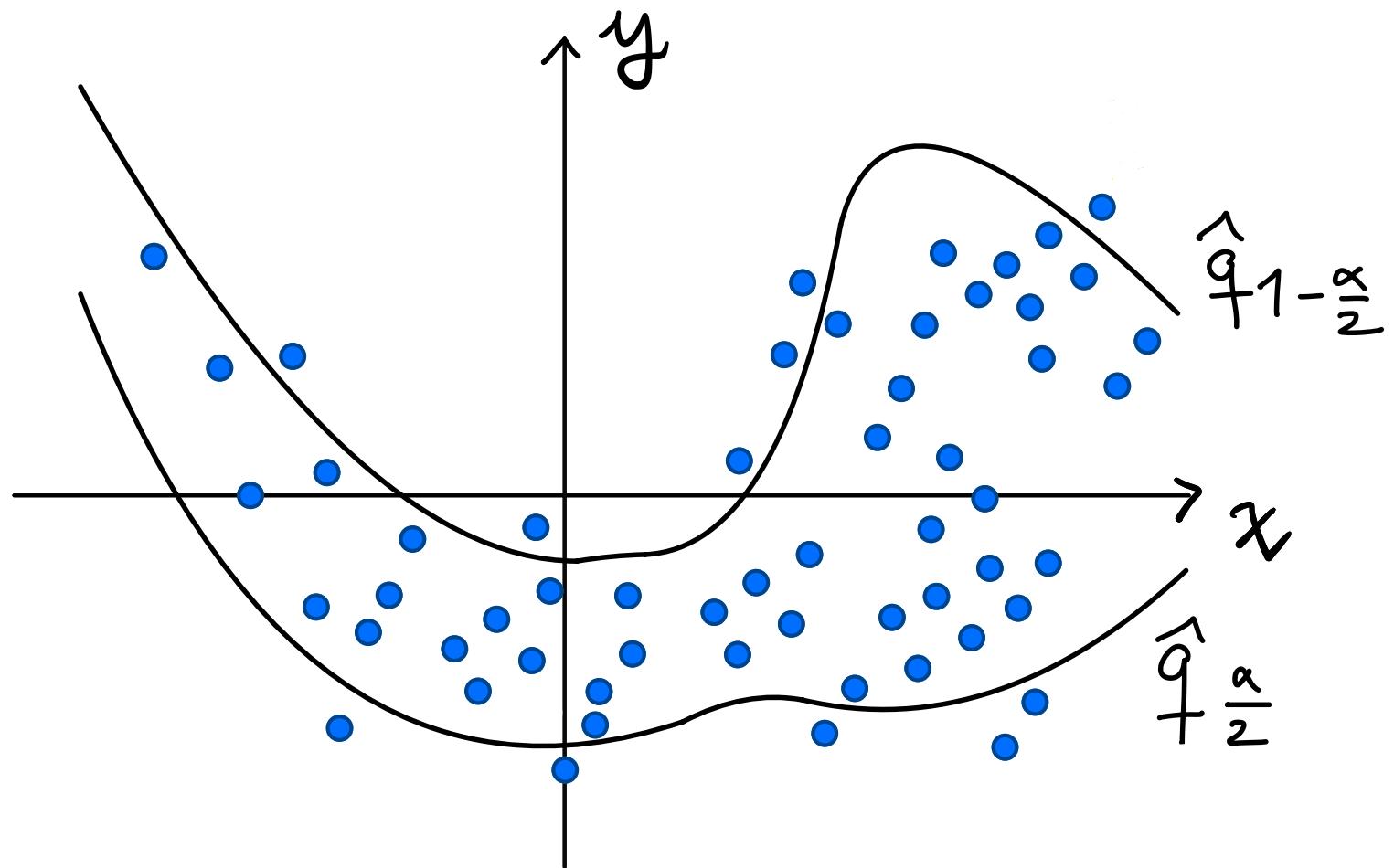
CQR : CONFORMAL

QUANTILE REGRESSION



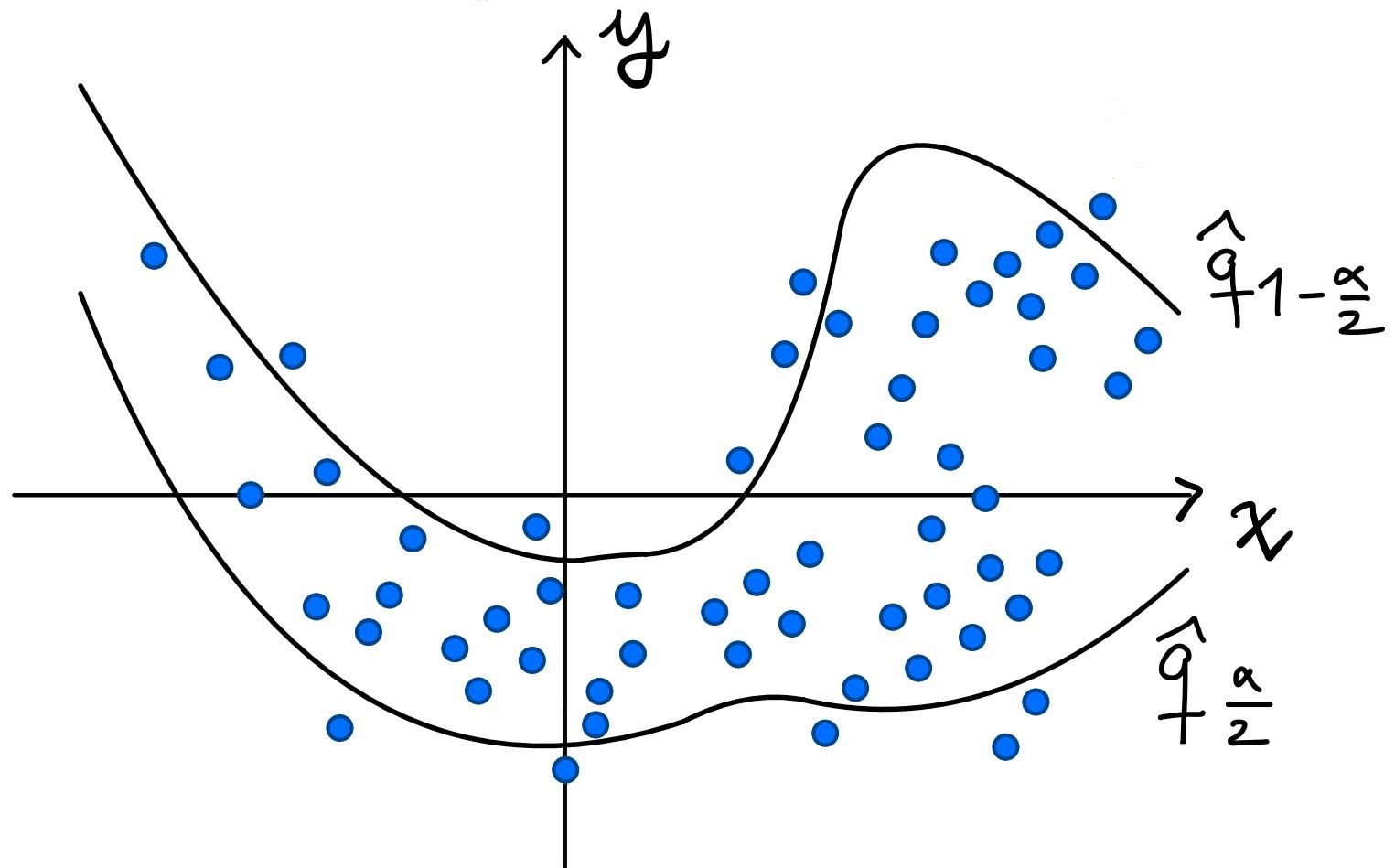
CQR

$$\hat{q}_t(x) \approx \mathbb{P}(Y < t | X = x)$$



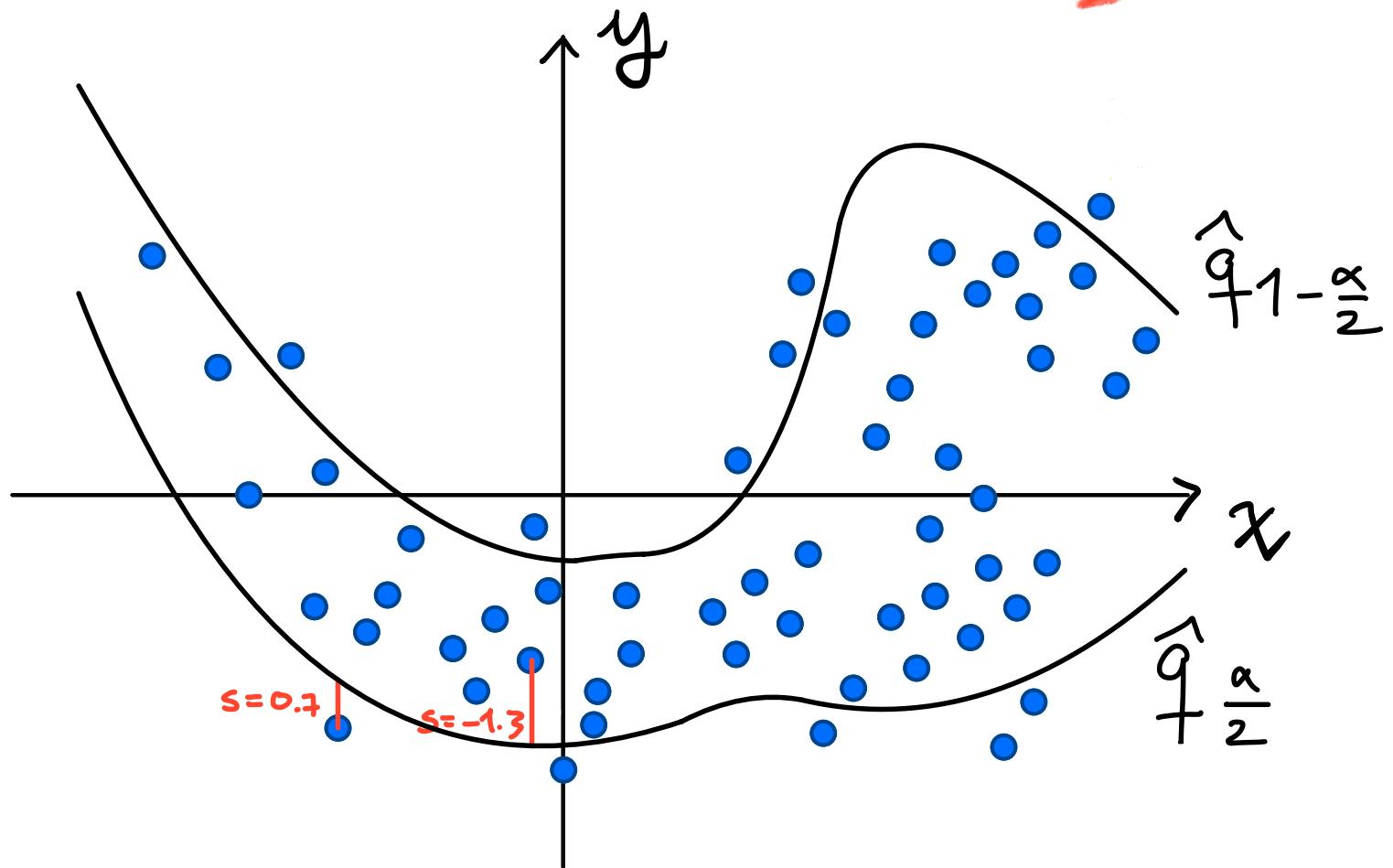
CQR

$$\mathbb{P}(Y \in [q_{\frac{\alpha}{2}}(x), q_{1-\frac{\alpha}{2}}(x)]) = ?$$



CQR : CALIBRATION

$$S_i = \max\{q_{\frac{\alpha}{2}}(x_i) - y_i, y_i + q_{1-\frac{\alpha}{2}}(x_i)\}$$



CQR : CALIBRATION

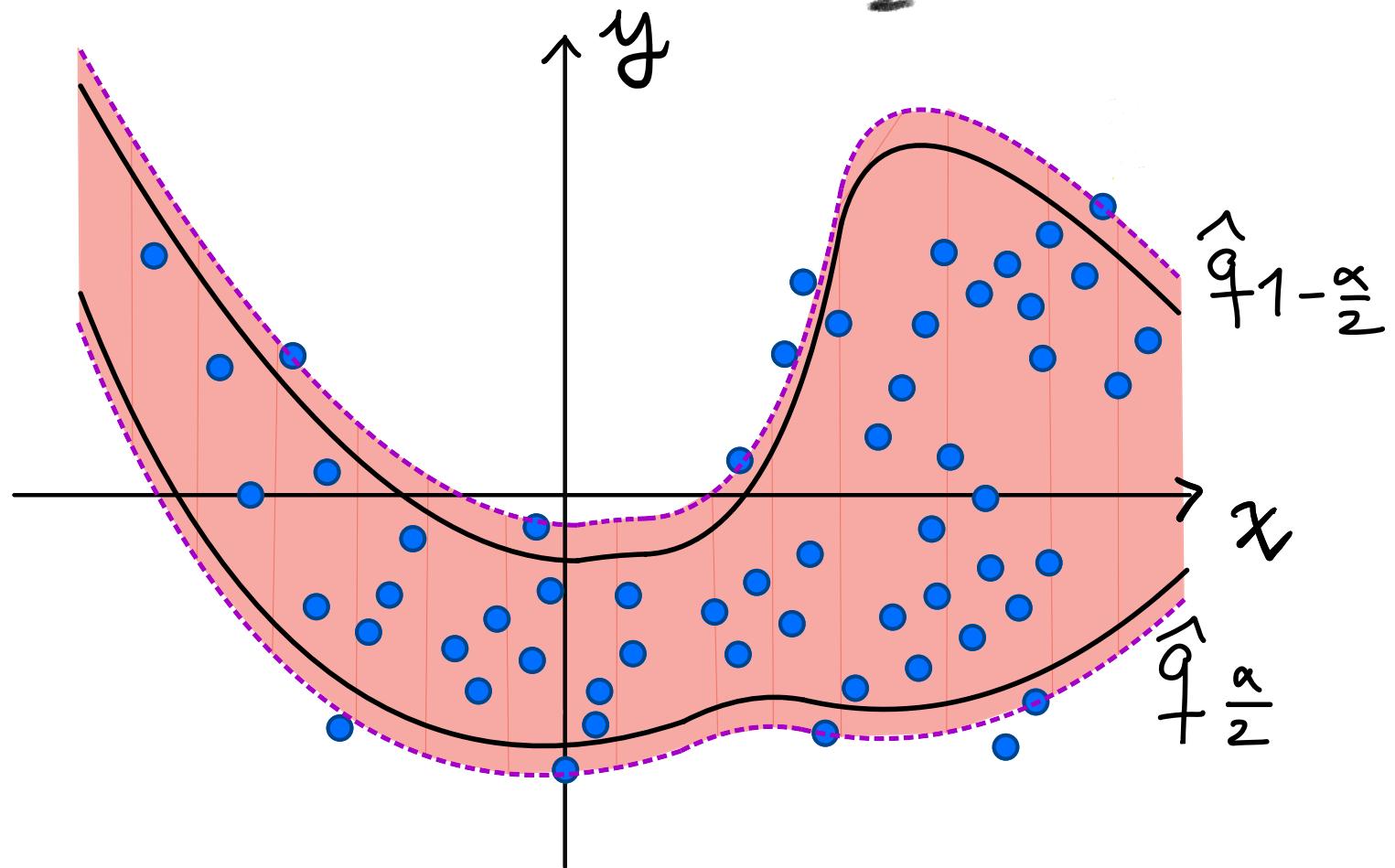
Calibration:

$S_\alpha := \text{the } \left\lceil \frac{(n+1)(1-\alpha)}{n} \right\rceil - \text{th}$

quantile of the scores S_1, \dots, S_n

CQR : CALIBRATION

$$C_\alpha(x) = \left[\hat{q}_{\frac{\alpha}{2}}(x) - \delta_\alpha, \hat{q}_{1-\frac{\alpha}{2}}(x) + \delta_\alpha \right]$$



GENERAL FORMULATION

Given : a predictor $\hat{f}: \mathcal{X} \rightarrow \mathcal{Y}$
a nominal error rate α
a scoring function $s: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}^+$
e.g. in regression $s(y, y') = |y - y'|$

Split conformal : Calibration set $\{(x_i, y_i)\}_{i=1}^n$

\Rightarrow non-conformity scores $R_i = s(\hat{Y}_i, y_i)$

$\Rightarrow q_\alpha = \left\lceil \frac{(1-\alpha)(n+1)}{n} \right\rceil$ -quantile of the $\{R_i\}_{i=1}^n$

$\Rightarrow C(\alpha) := \{y: s(\hat{y}, y) \leq q_\alpha\}$

CONFORMAL CLASSIFICATION

- Least Ambiguous Set-Valued Classifiers (LAC)
- Adaptive Prediction Sets (APS)
- Regularized Adaptive Prediction Sets (RAPS)

LAC

Given : $\hat{\pi}$ softmax classifier

α nominal error rate

Use the nonconformity scores

$$s(\hat{\pi}, y) = 1 - \underbrace{\hat{\pi}_y}_{\text{in purple}}$$

i.e. 1- the probability
of the true label.

APS

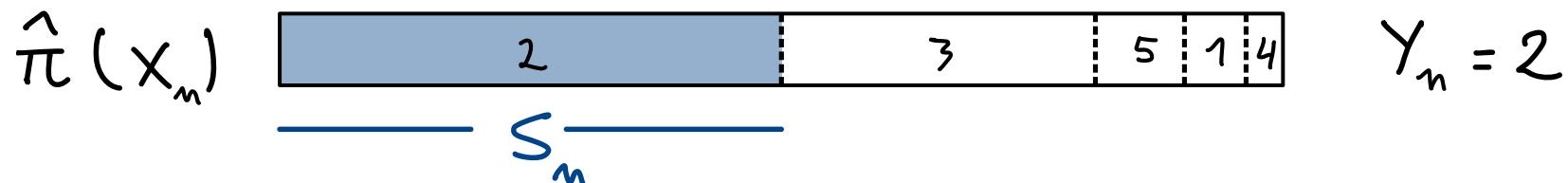
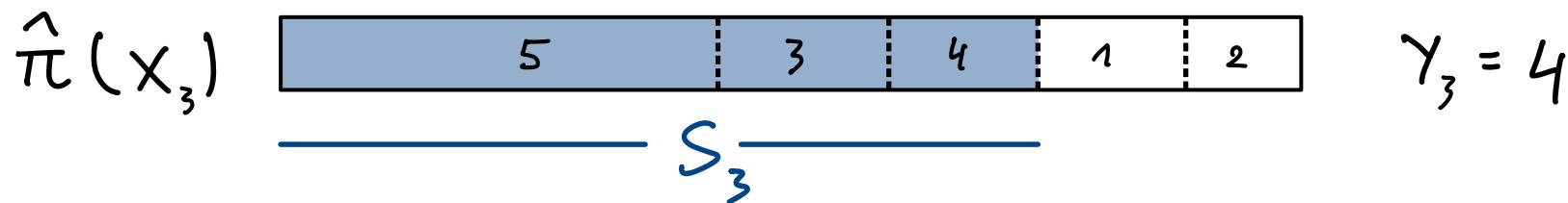
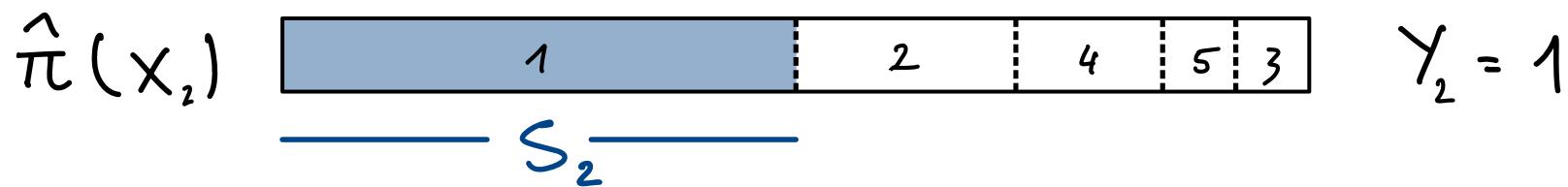
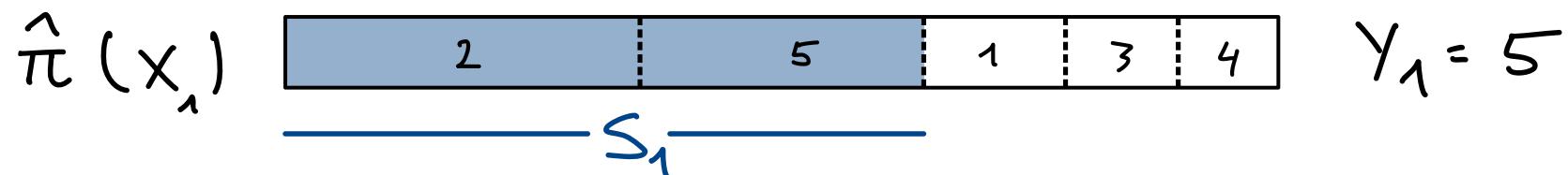
Given : $\hat{\pi}$ softmax classifier
 α nominal error rate

Rank: $\hat{\pi}_{(1)}(x) \geq \hat{\pi}_{(2)}(x) \geq \dots \geq \hat{\pi}_{(k)}(x)$

$$L(x, \hat{\pi}, z) := \min_{c \in \{1, \dots, k\}} \left\{ \hat{\pi}_{(1)}(x) + \dots + \hat{\pi}_c(x) \geq z \right\}$$

APS

Calibration:



APS

Calibration:

$S_\alpha := \text{the } \left\lceil \frac{(n+1)(1-\alpha)}{n} \right\rceil - \text{th}$

quantile of the scores S_1, \dots, S_n

APS

Inference: $\hat{C}_\alpha(x) = \{(1), (2), \dots, (K)\}$

where $K = \min \{i : \hat{\pi}_{(1)} + \dots + \hat{\pi}_{(i)} \geq \delta_\alpha\}$



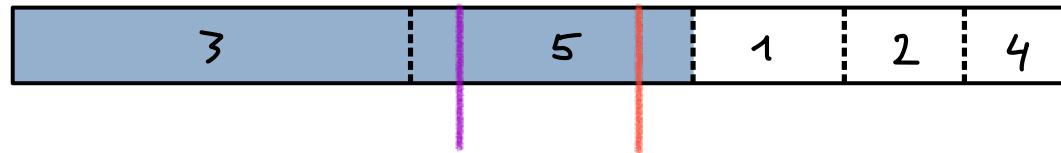
$$\delta_\alpha \Rightarrow C_\alpha(x) = \{3, 5\}$$

Guarantee:

$$P(Y_{n+1} \in \hat{C}_\alpha(X_{n+1})) \geq 1 - \alpha$$

APS WITH RANDOMIZATION

$\hat{\pi}(x)$



s_α $u \leftarrow$ Uniform random variable in

$$[\pi_3, \pi_3 + \pi_5]$$

$$\Rightarrow \hat{c}_\alpha(x) := \begin{cases} \{3\} & \text{if } u \leq s_\alpha \\ \{3, 5\} & \text{if } u > s_\alpha \end{cases}$$



Same guarantee, tighter prediction sets!

RAPS

Given : $\hat{\pi}$ softmax classifier

α nominal error rate

Rank: $\hat{\pi}_{(1)}(x) \geq \hat{\pi}_{(2)}(x) \geq \dots \geq \hat{\pi}_{(K)}(x)$

$$S_i := \hat{\pi}_{(1)} + \dots + \hat{\pi}_{(k)} + \lambda (K - K_{\text{reg}} + 1)$$

K such that $(k) = y_i$

rank of y_i

hyper-parameters

RAPS

Prediction set:

$$\hat{\mathcal{C}}_a(x) = \{1, \dots, (k)\}$$

where

$$K = \max \{ i : \pi_{(1)} + \dots + \pi_{(i)} + \lambda(i - K_{\text{reg}} + 1) \leq S_a \} + 1$$



More stable than APS if softmax is "noisy" i.e. many classes w. low proba.