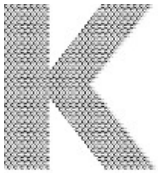
Detection of Classifier Inconsistencies in Image Steganalysis

Daniel Lerch-Hostalot & David Megías

7th ACM Workshop on Information Hiding and Multimedia Security
Paris, France July 3-5

Universitat Oberta de Catalunya
Internet Interdisciplinary Institute (IN3)



Cryptography and Information Security for Open Networks

Outline

1. Introduction
2. Training and testing sets
3. Detection of inconsistencies
4. Prediction of the classifier's error
5. Experimental results
6. Conclusions and future work

Outline

1. **Introduction**
2. Training and testing sets
3. Detection of inconsistencies
4. Prediction of the classifier's error
5. Experimental results
6. Conclusions and future work

Introduction

SCENARIO

- Batch Steganography & Pooled Steganalysis
- Attack to known algorithm and bit rate

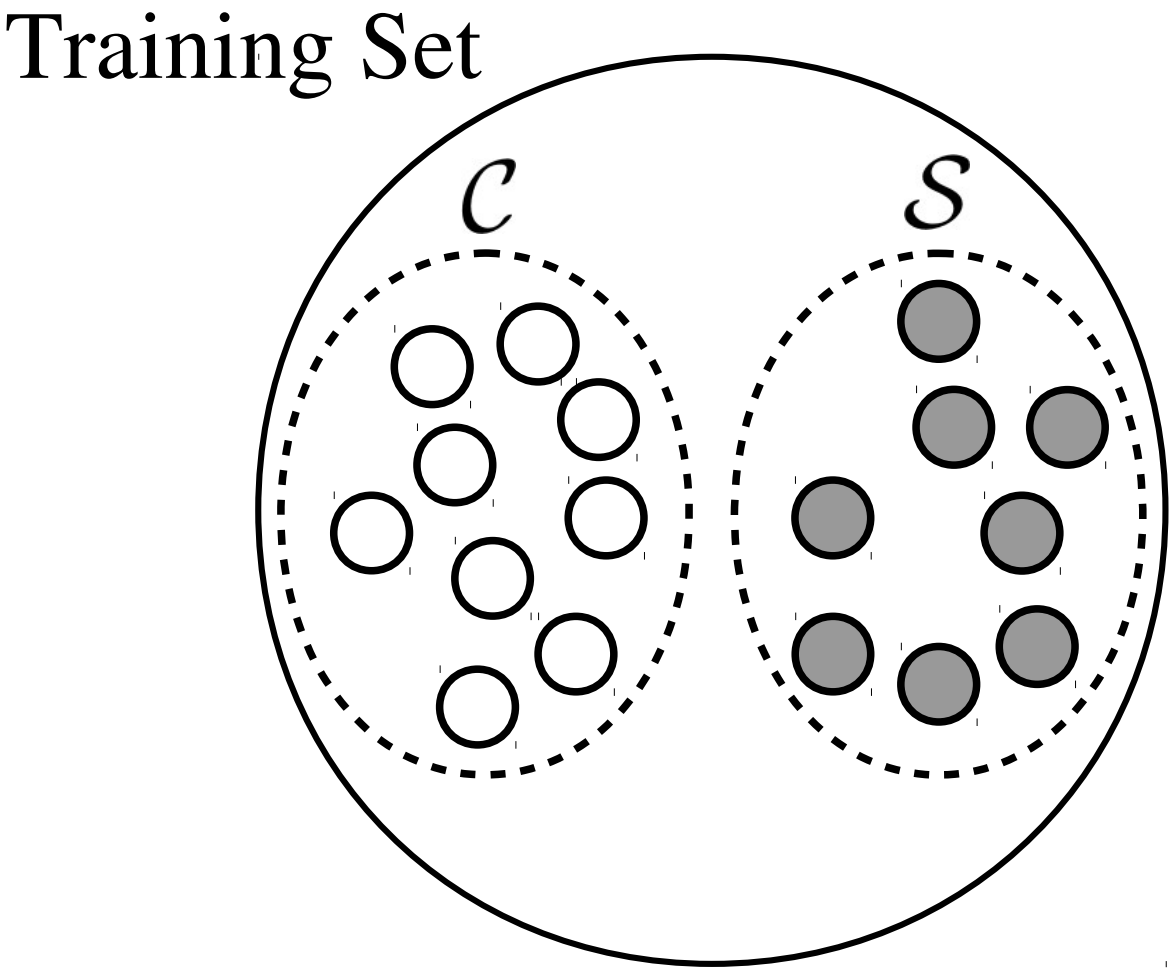
PROPOSED METHOD

- Detection of inconsistencies in classification
- Prediction of the classifier's error

Outline

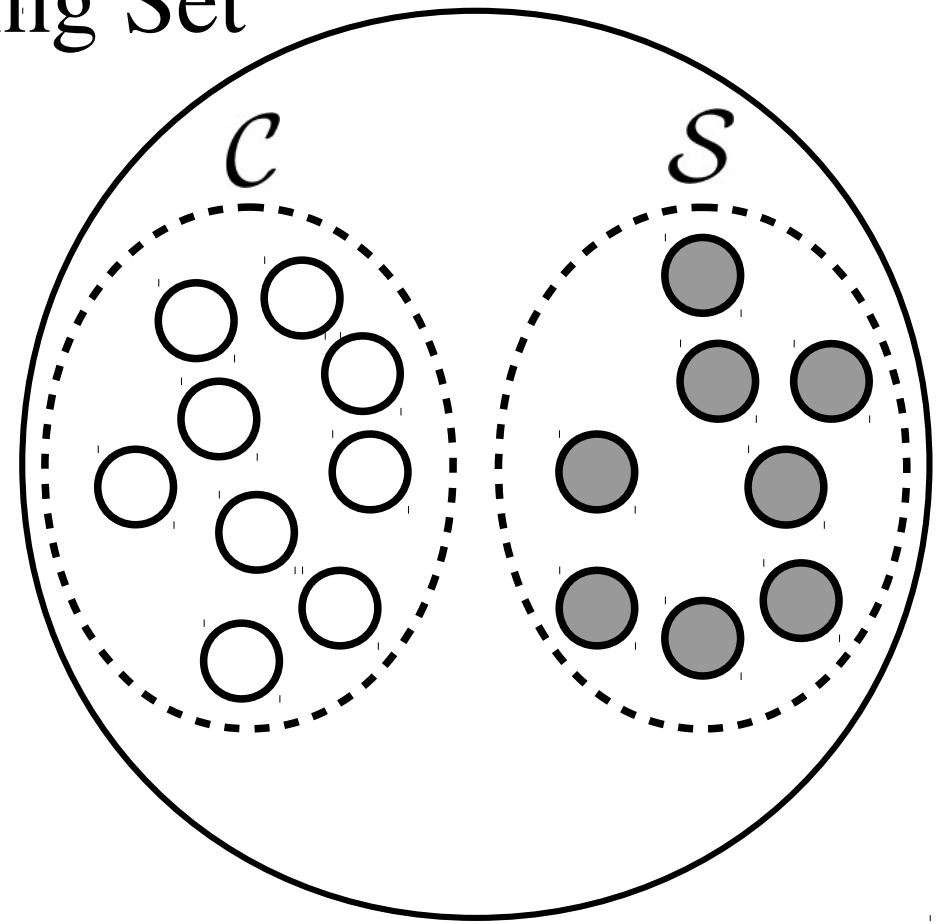
1. Introduction
2. **Training and testing sets**
3. Detection of inconsistencies
4. Prediction of the classifier's error
5. Experimental results
6. Conclusions and future work

Training



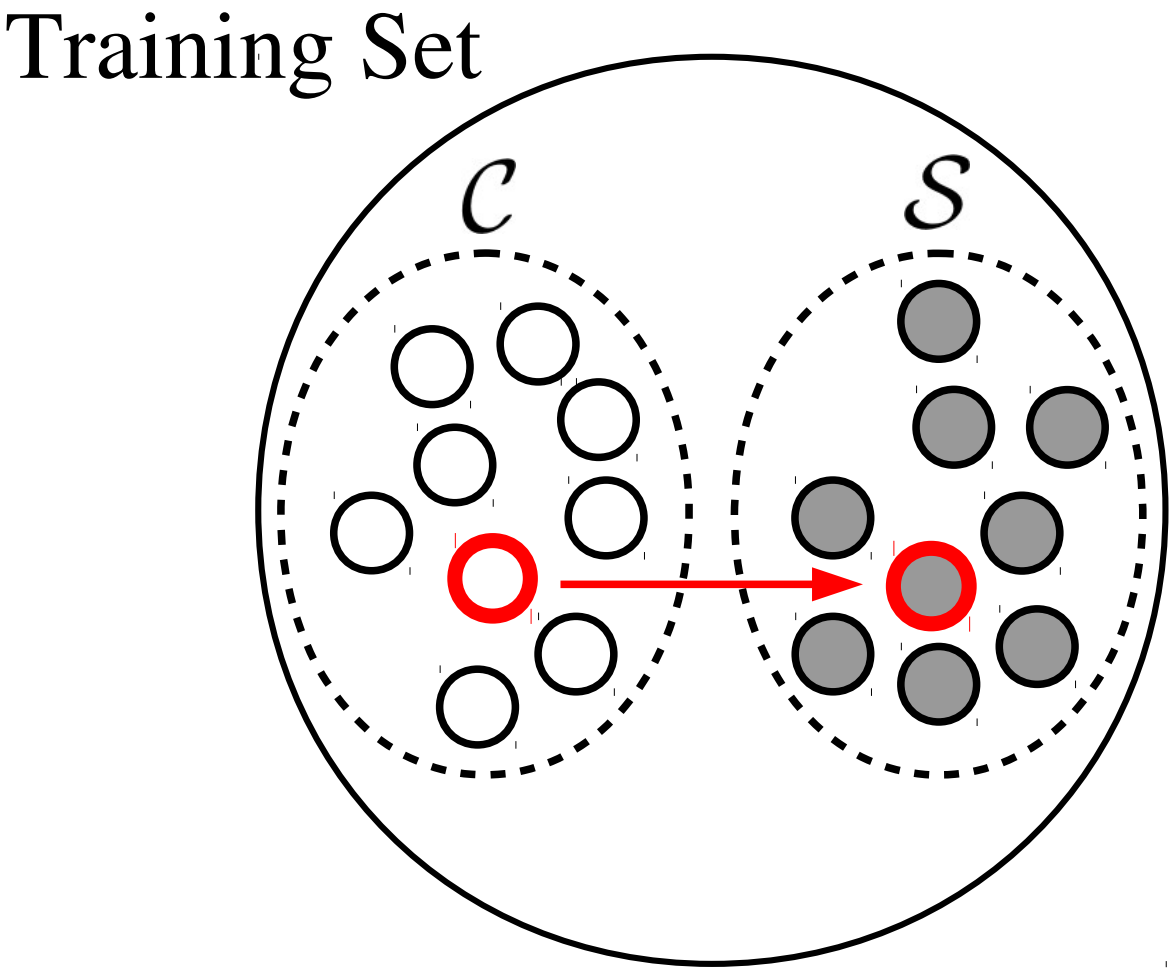
Training

Training Set

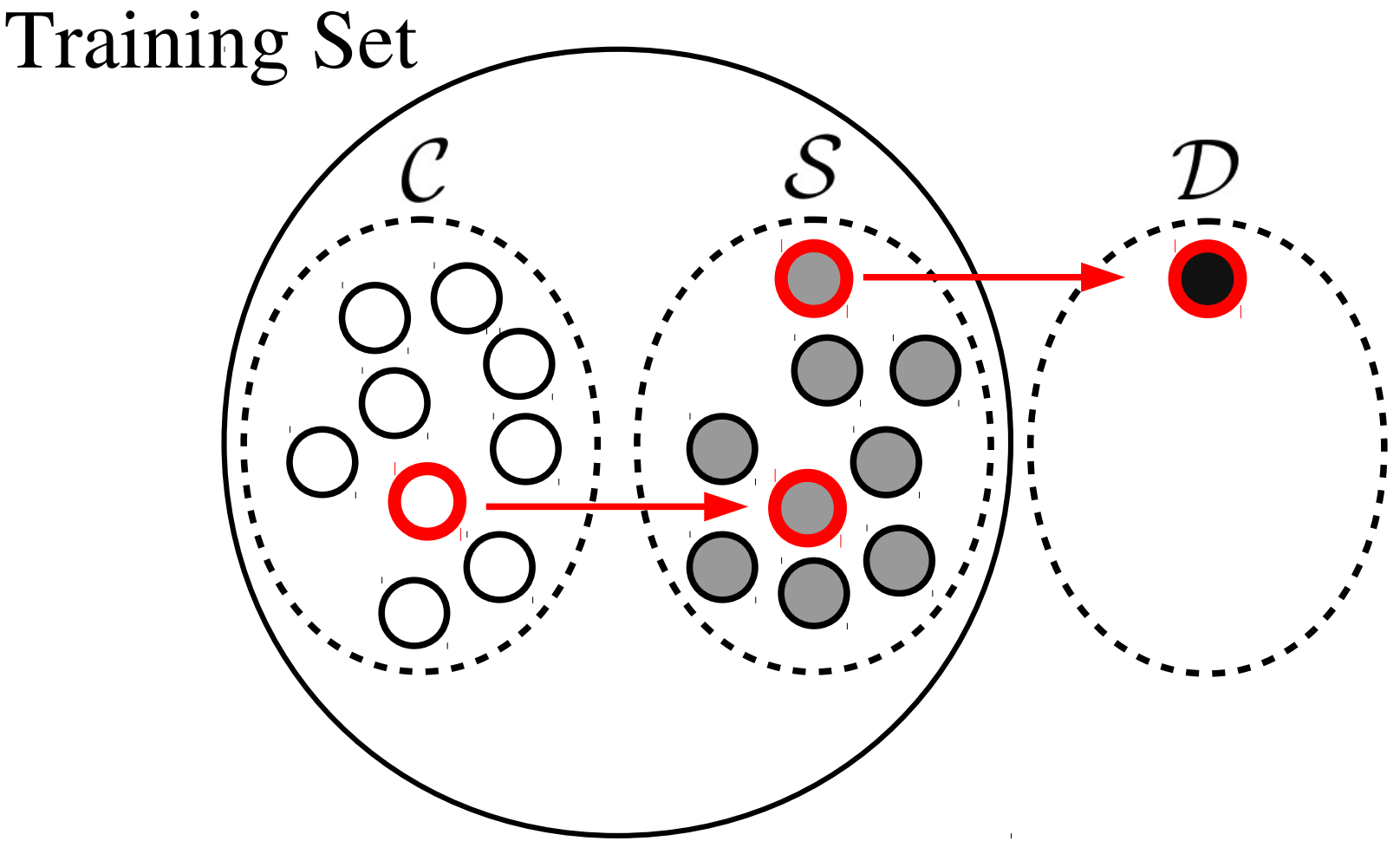


$$\hat{f} : X \longmapsto \{\mathcal{C}, S\}$$

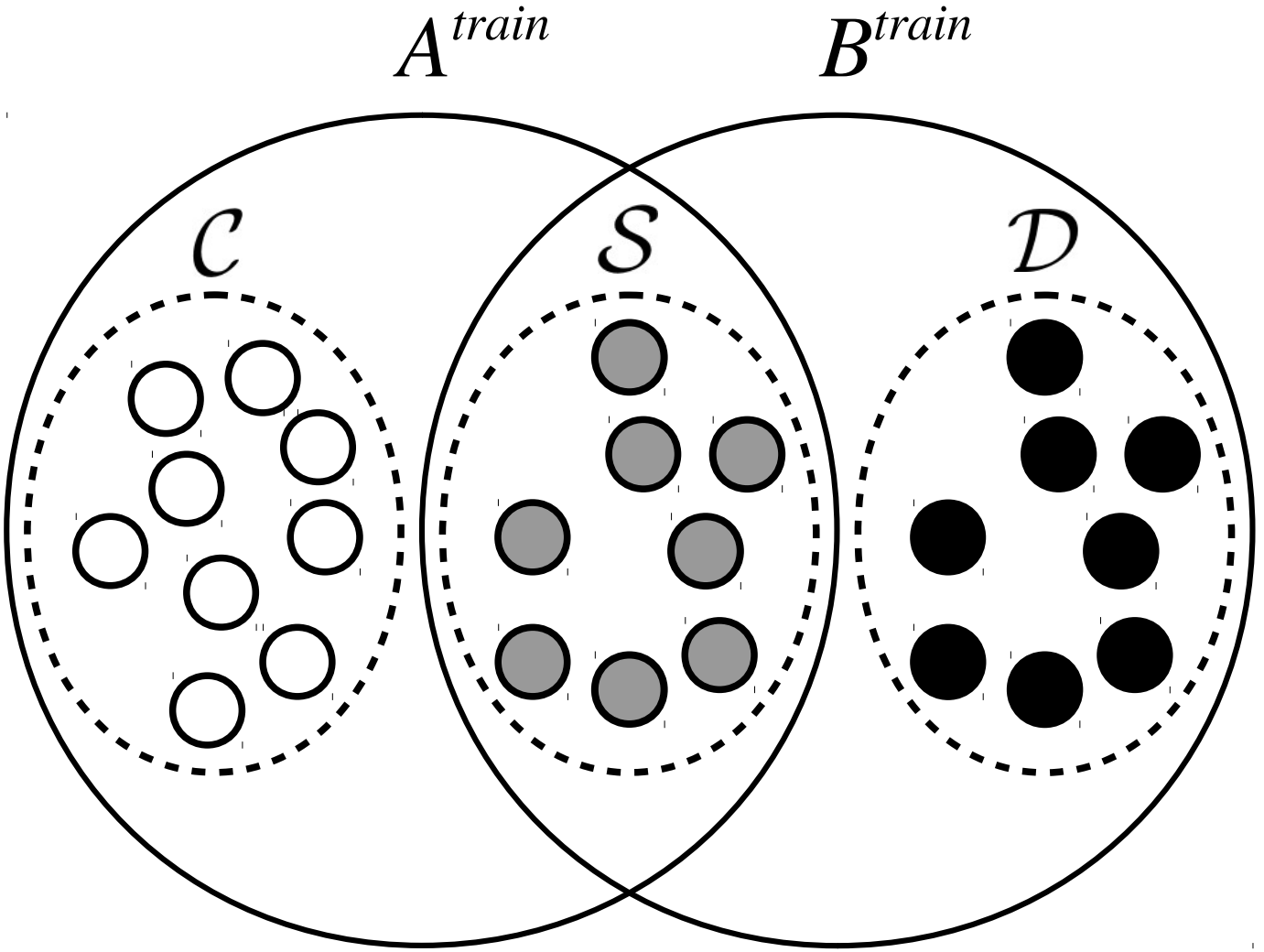
Training



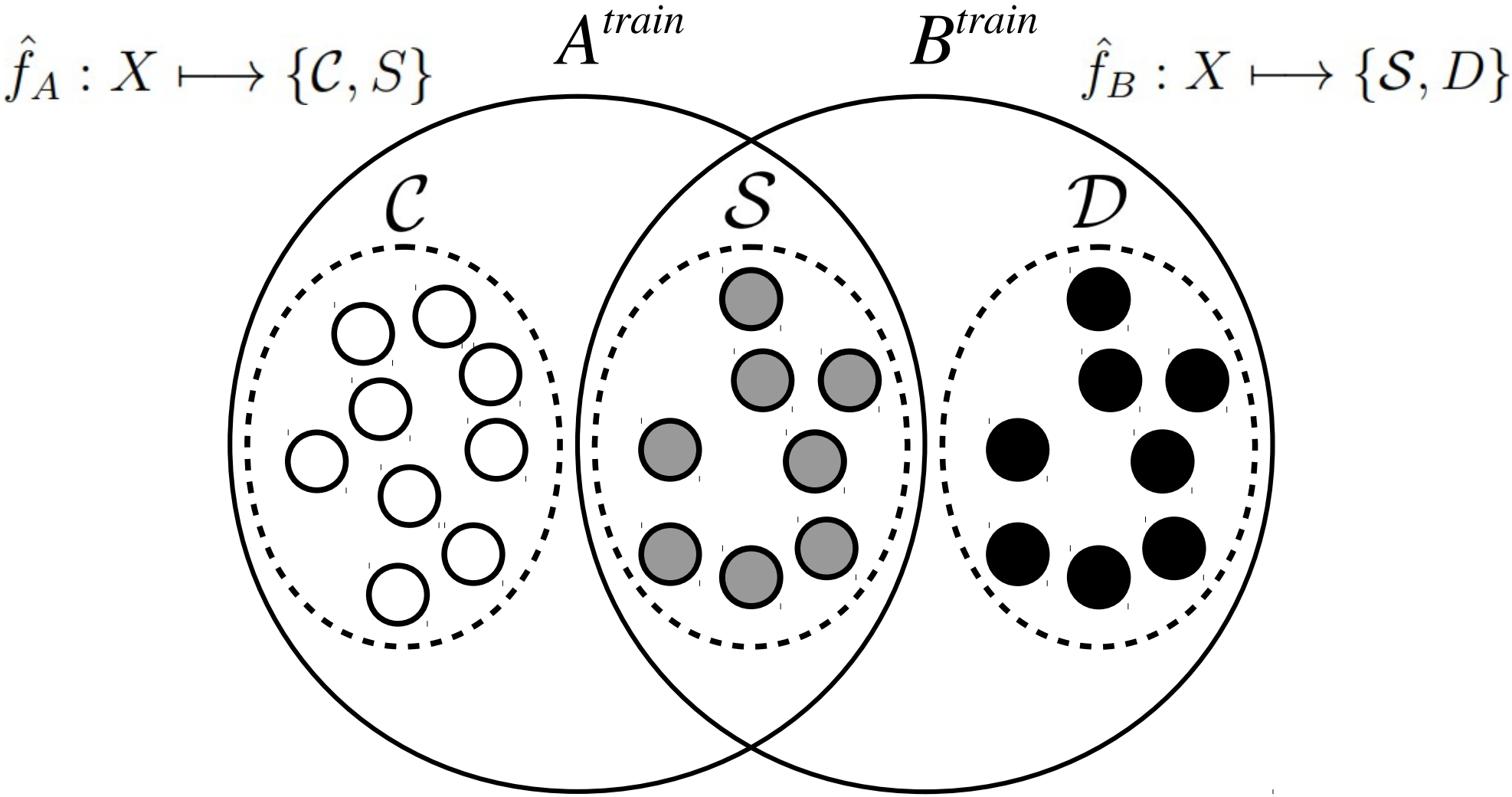
Training



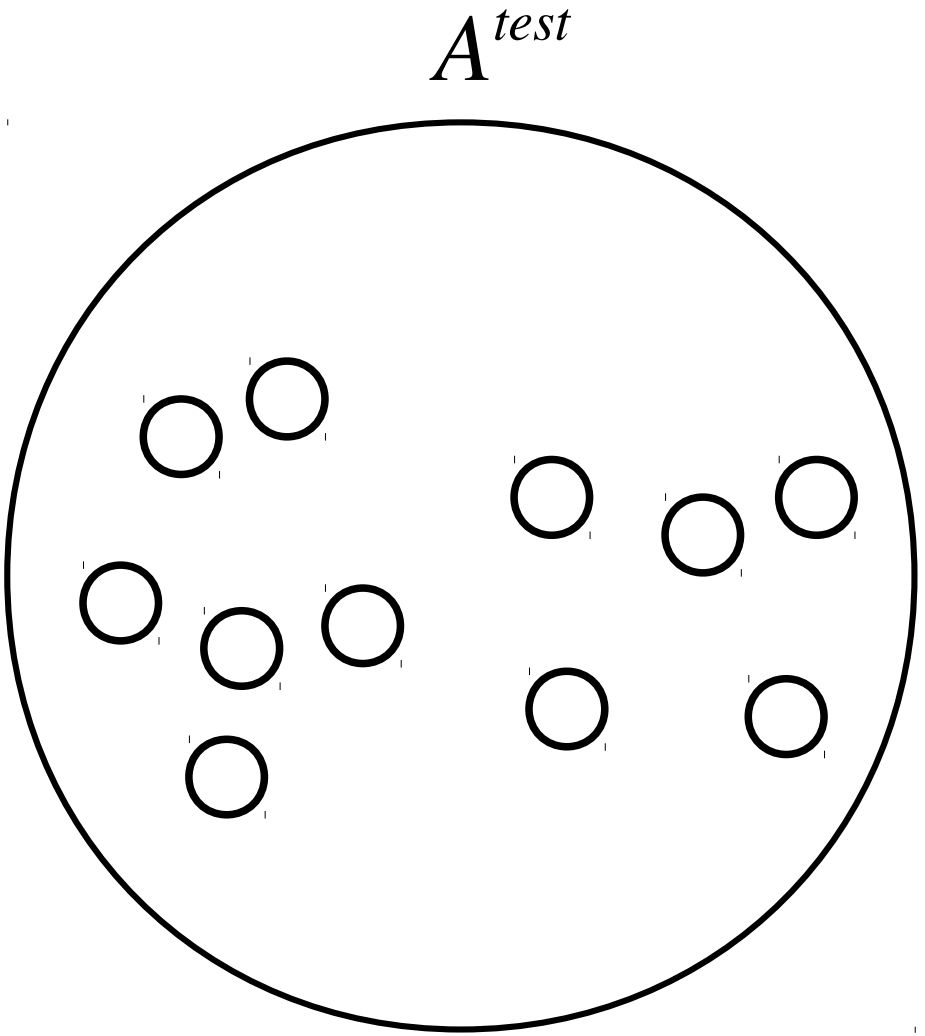
Training



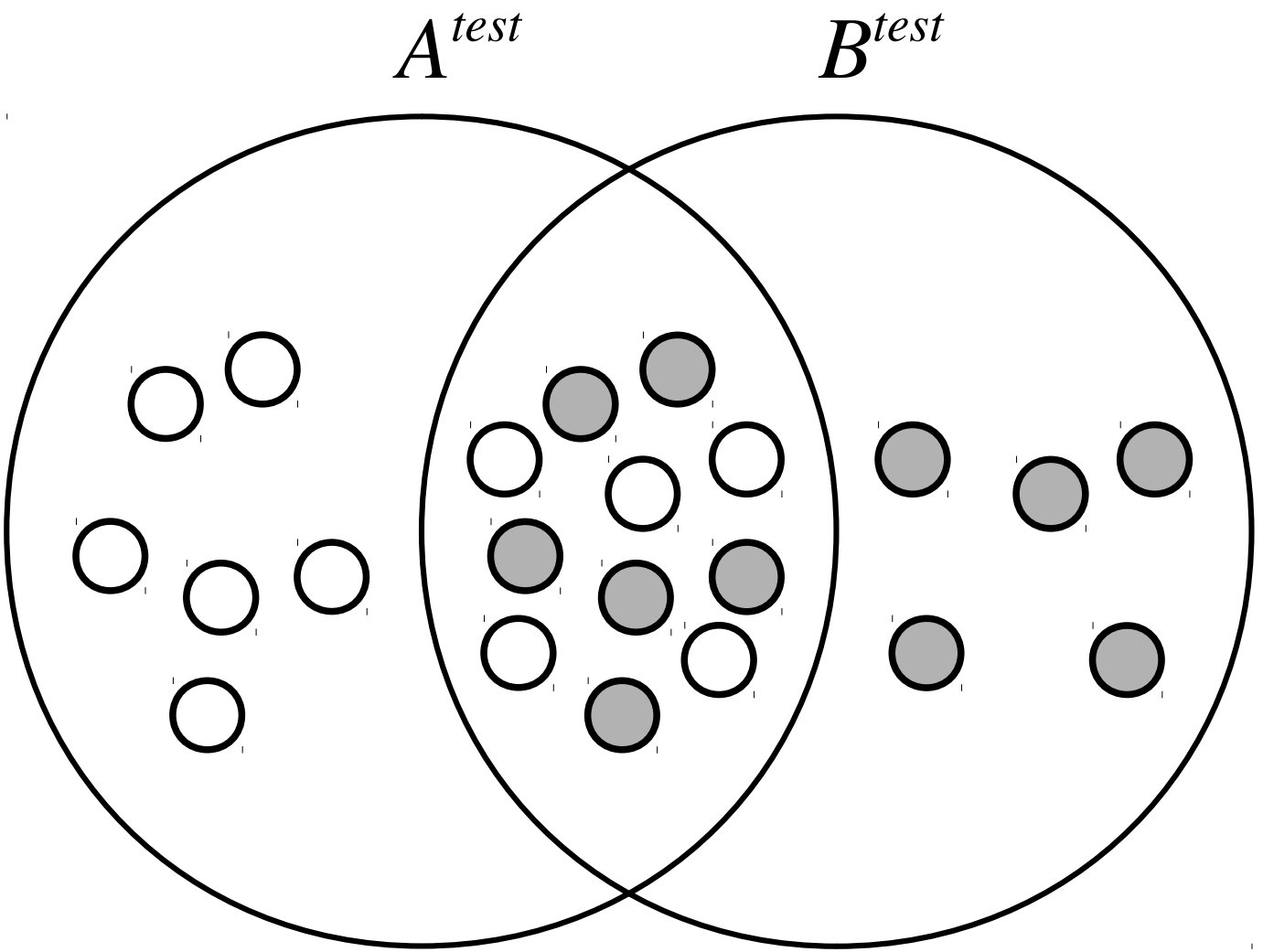
Training


$\hat{f}_A : X \longmapsto \{\mathcal{C}, S\}$
A^{train}
B^{train}
$\hat{f}_B : X \longmapsto \{\mathcal{S}, D\}$
$\mathcal{C}$
$\mathcal{S}$
$\mathcal{D}$

Testing Sets



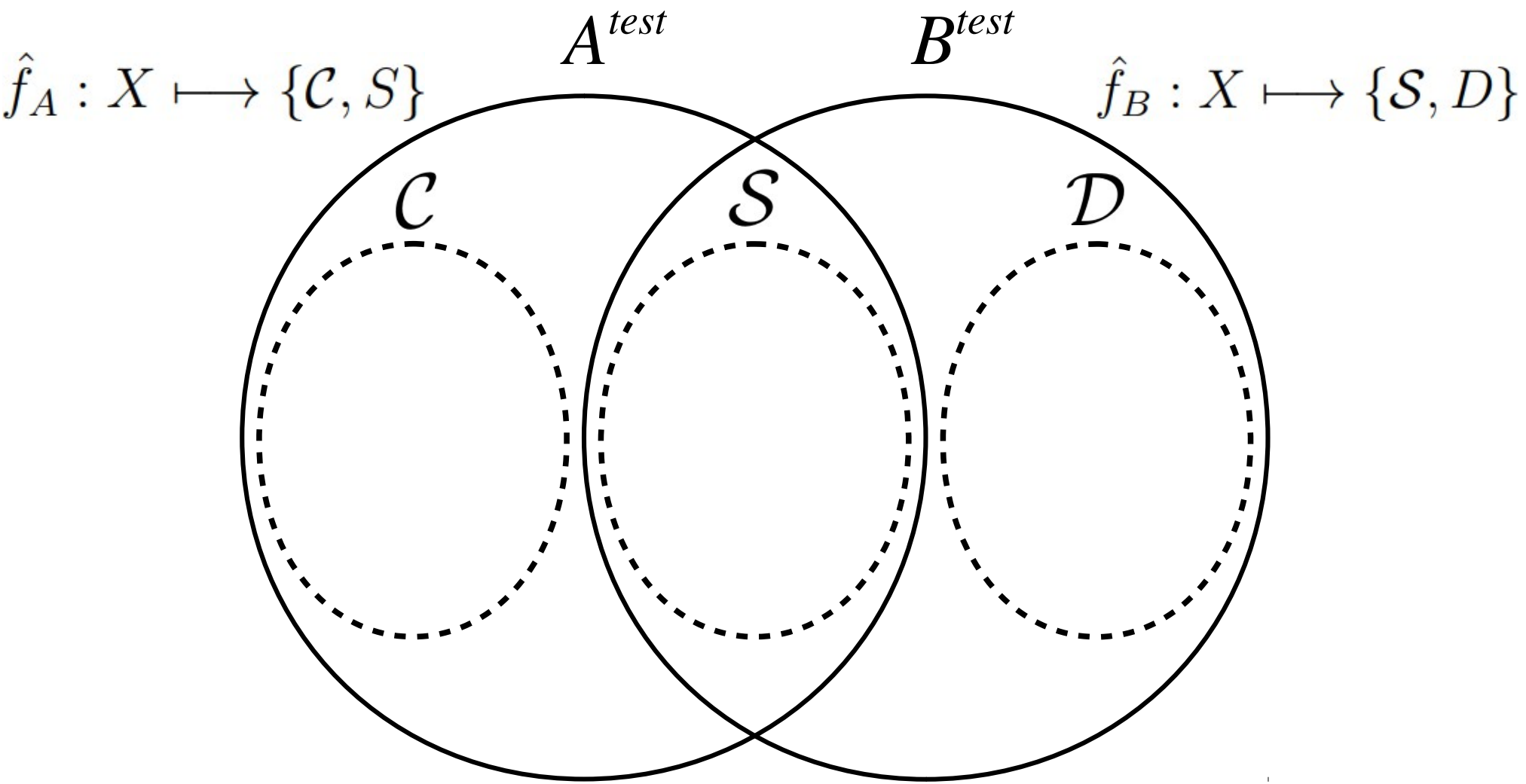
Testing Sets



Outline

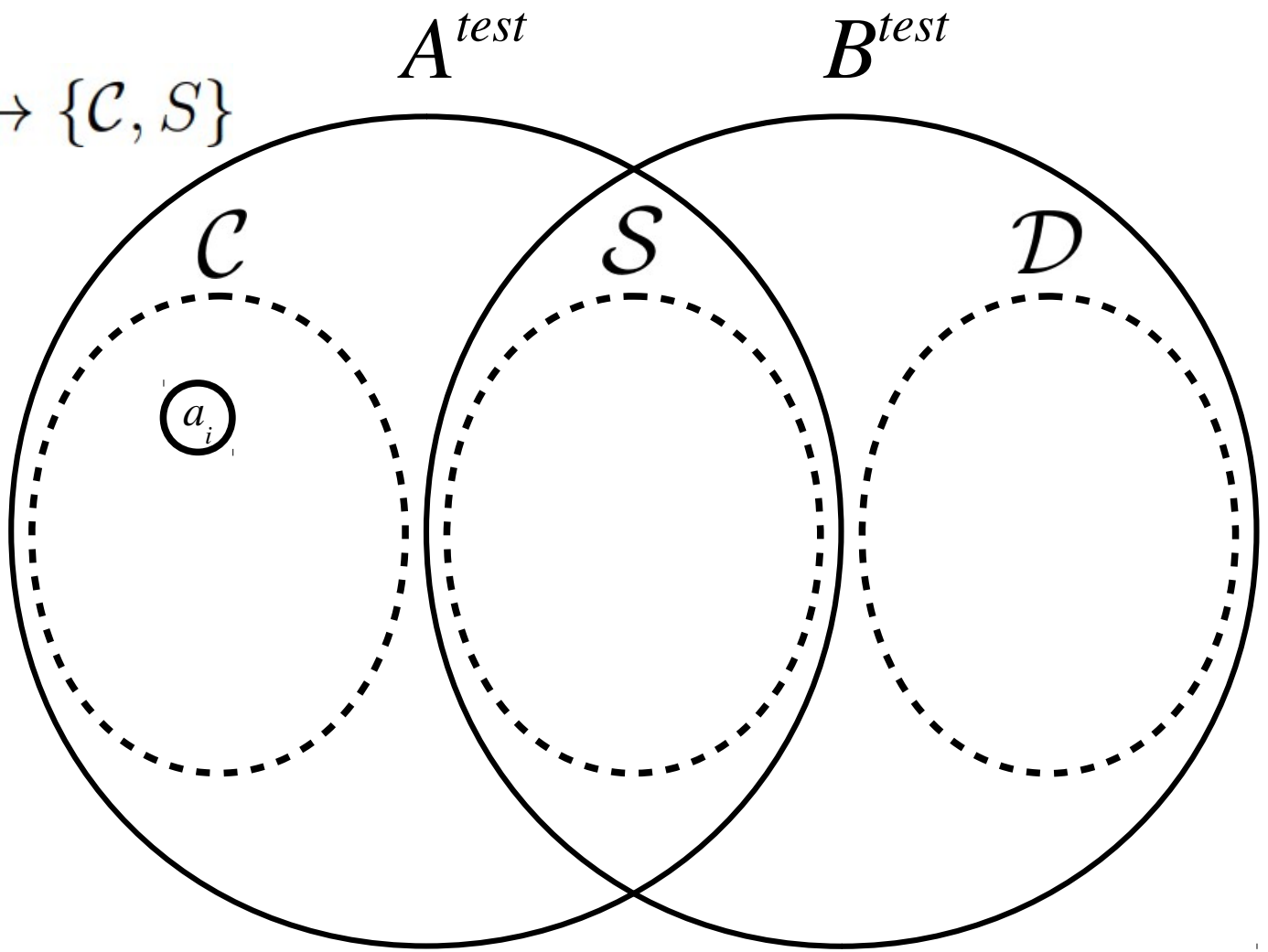
1. Introduction
2. Training and testing sets
3. **Detection of inconsistencies**
4. Prediction of the classifier's error
5. Experimental results
6. Conclusions and future work

Consistent Samples

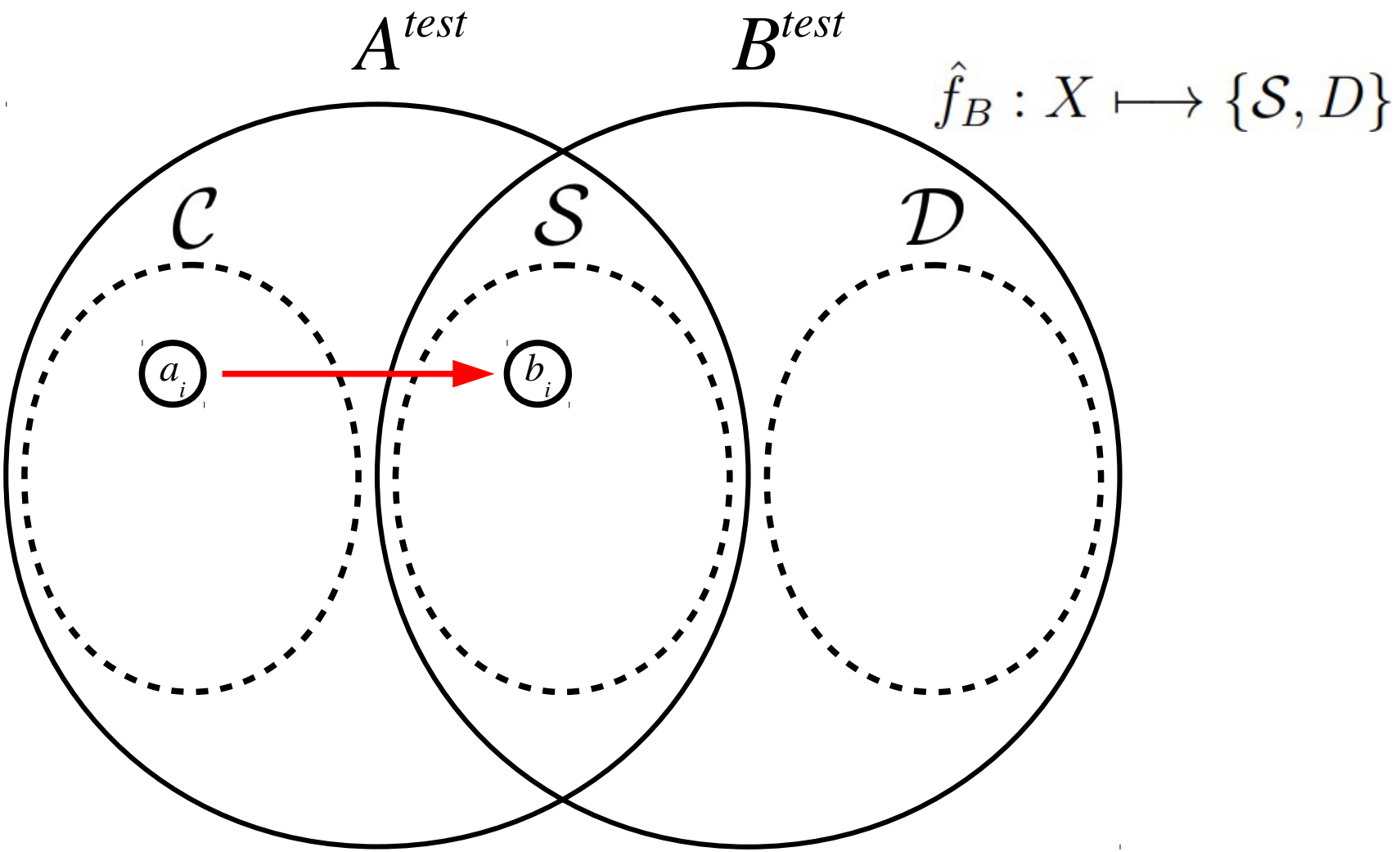


Consistent Samples

$\hat{f}_A : X \longmapsto \{\mathcal{C}, \mathcal{S}\}$

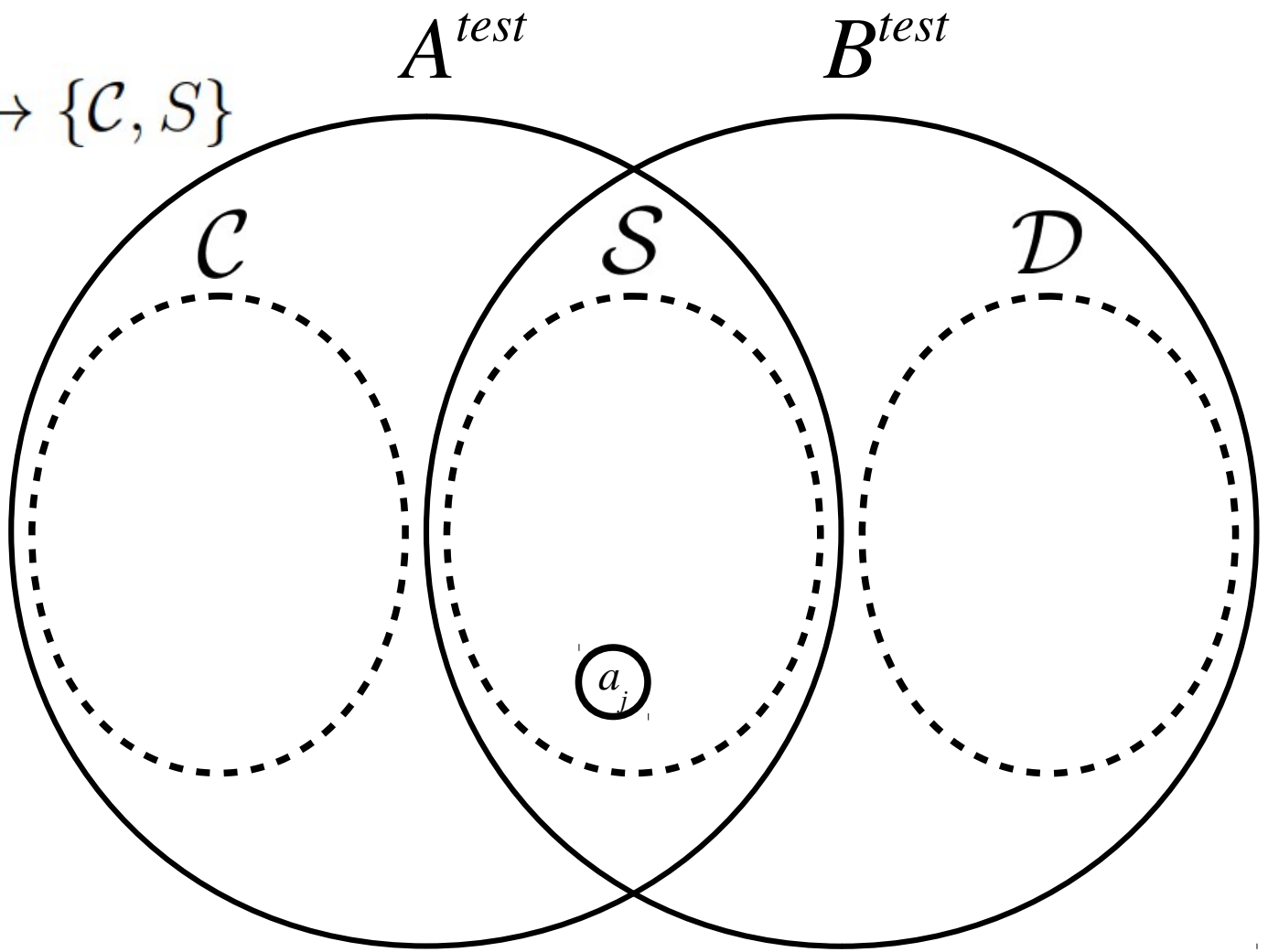


Consistent Samples

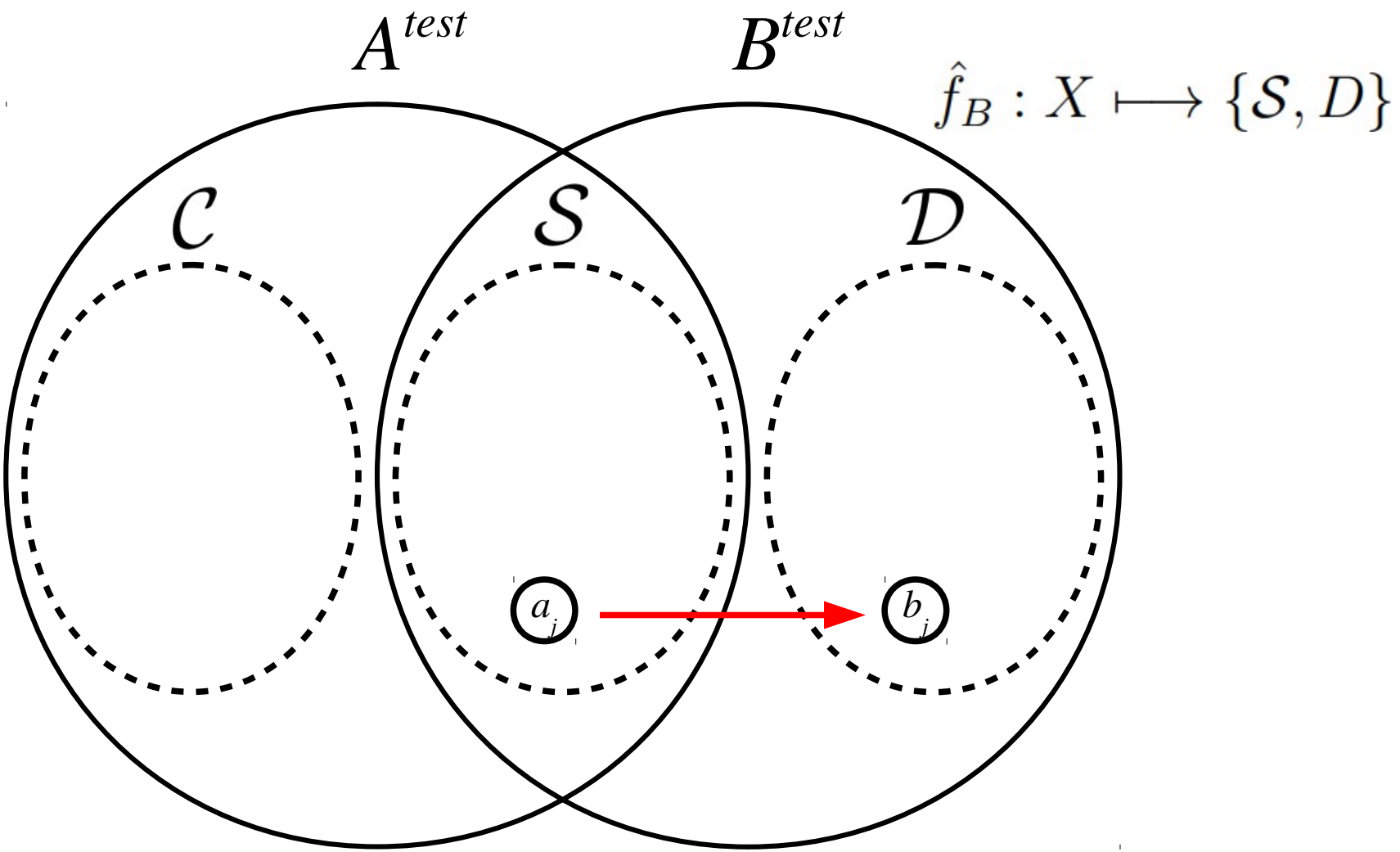


Consistent Samples

$\hat{f}_A : X \longmapsto \{\mathcal{C}, S\}$



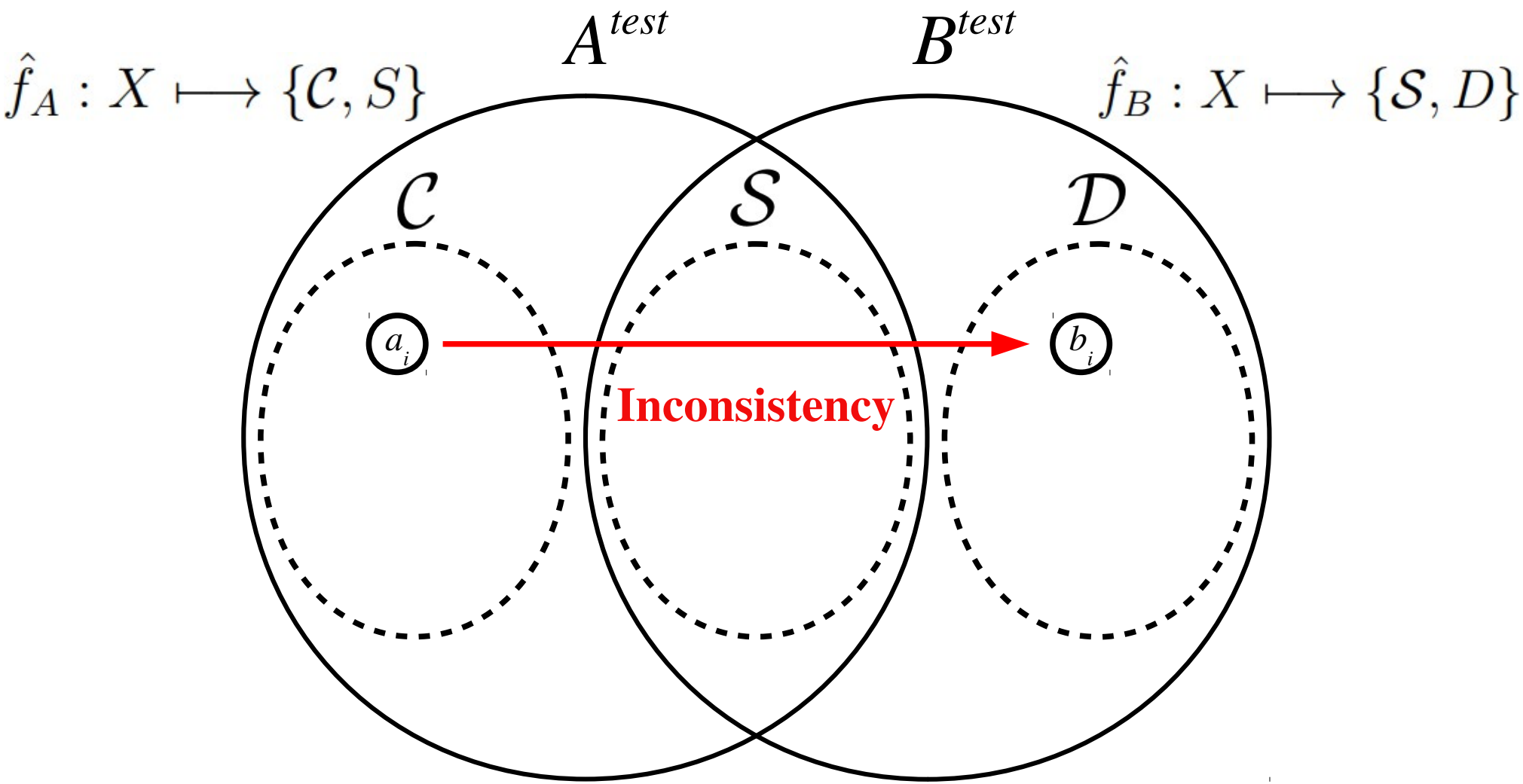
Consistent Samples



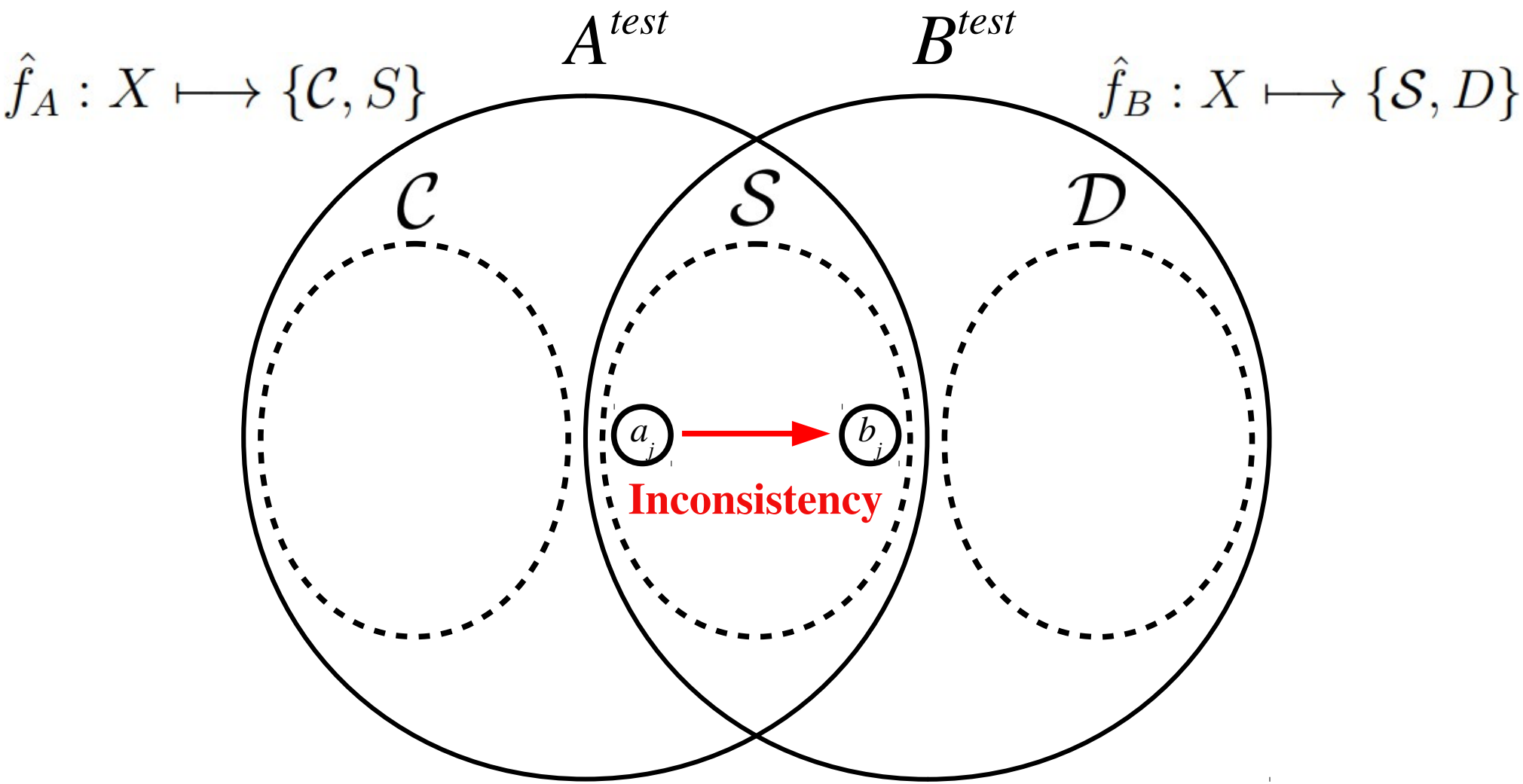
F1 Inconsistencies

$$F_1(i) \equiv \begin{cases} \text{If } \hat{f}_A(a_i) = \mathcal{S}_A, & \text{If } (\hat{f}_B(b_i) \neq \mathcal{D}_B) \\ & \text{then output “inconsistency”,} \\ \text{Otherwise,} & \text{If } (\hat{f}_B(b_i) \neq \mathcal{S}_B) \\ & \text{then output “inconsistency”.} \end{cases}$$

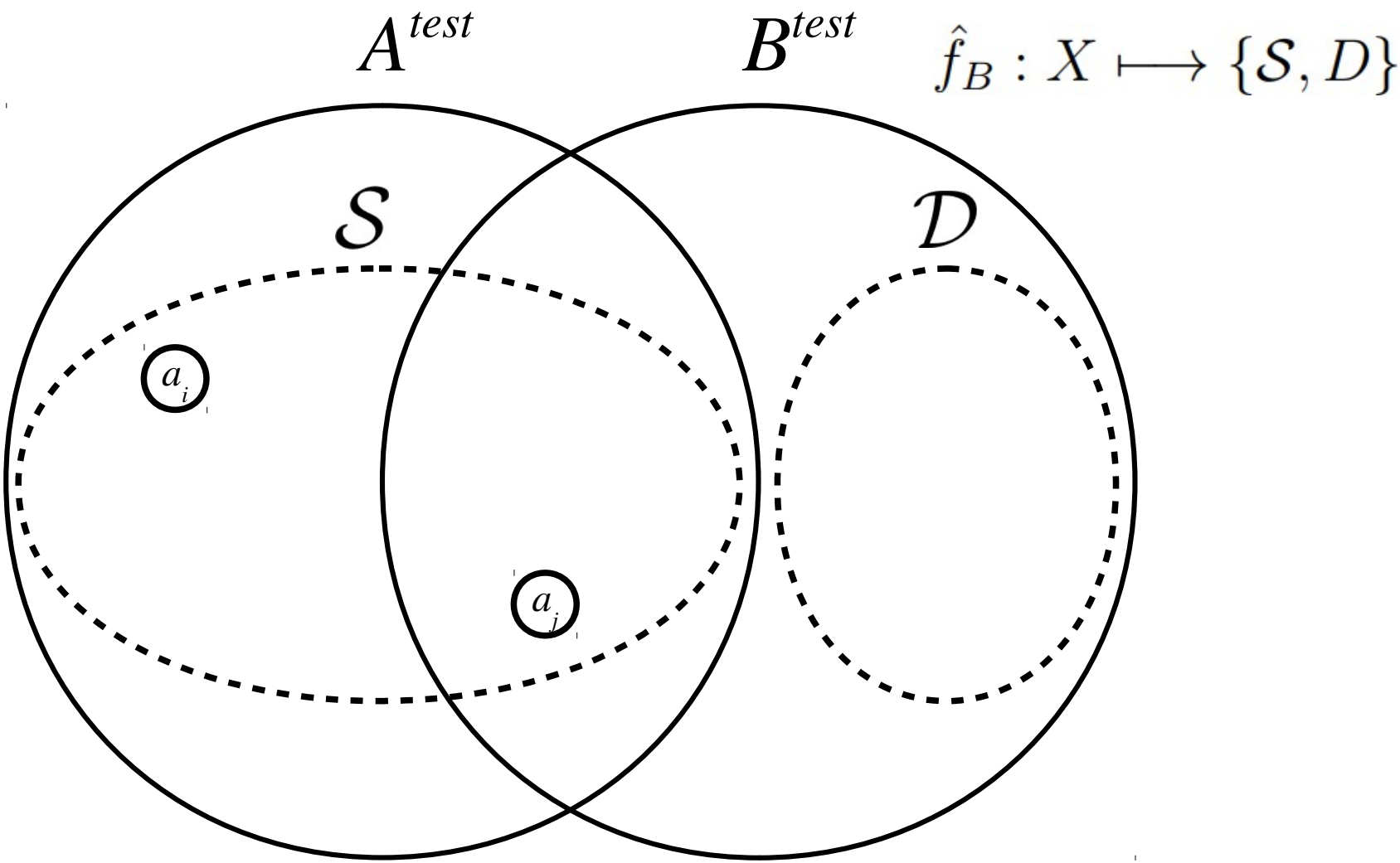
F1 Inconsistencies



F1 Inconsistencies

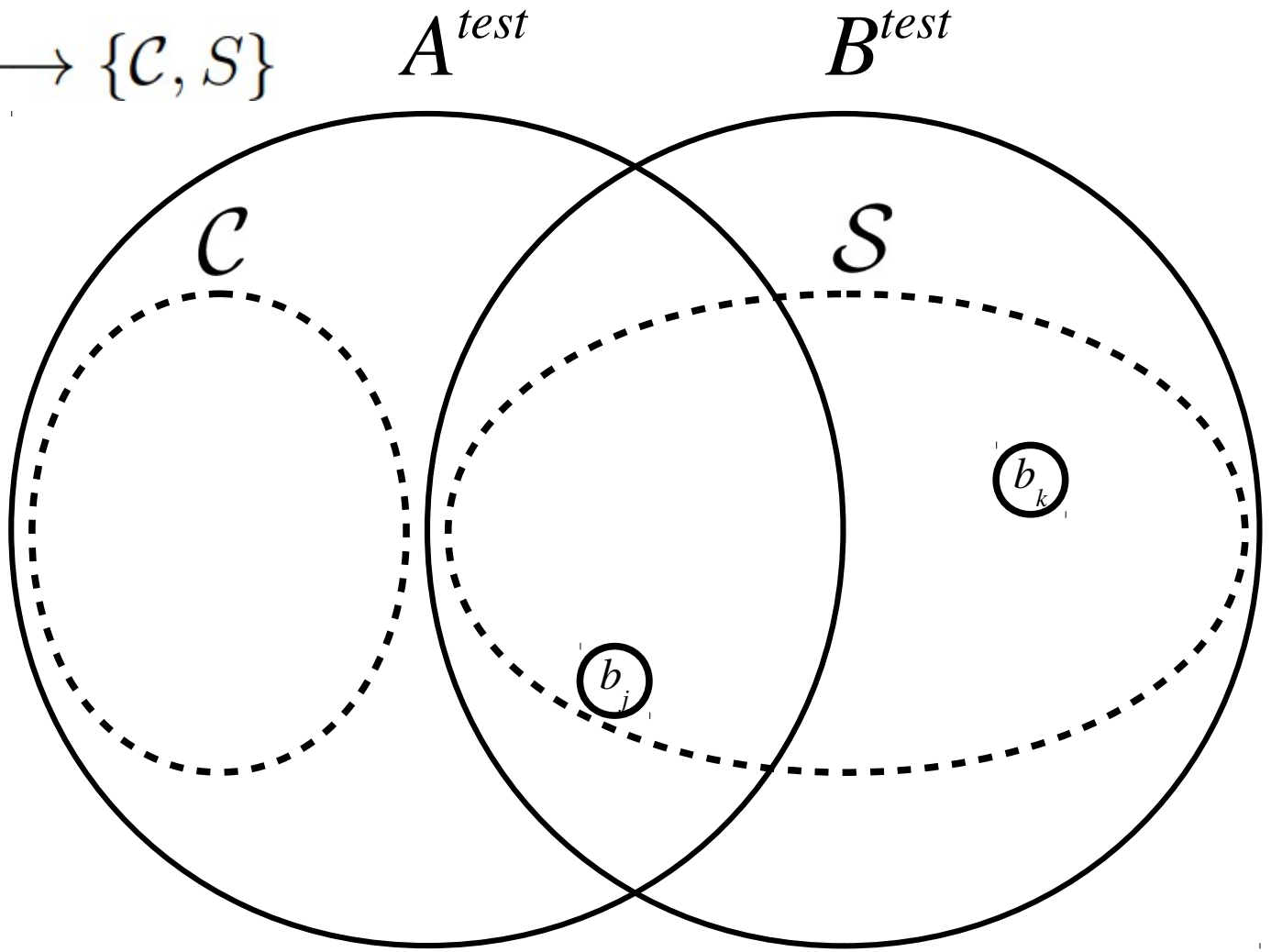


Consistent Samples 2



Consistent Samples 2

$\hat{f}_A : X \longmapsto \{\mathcal{C}, S\}$

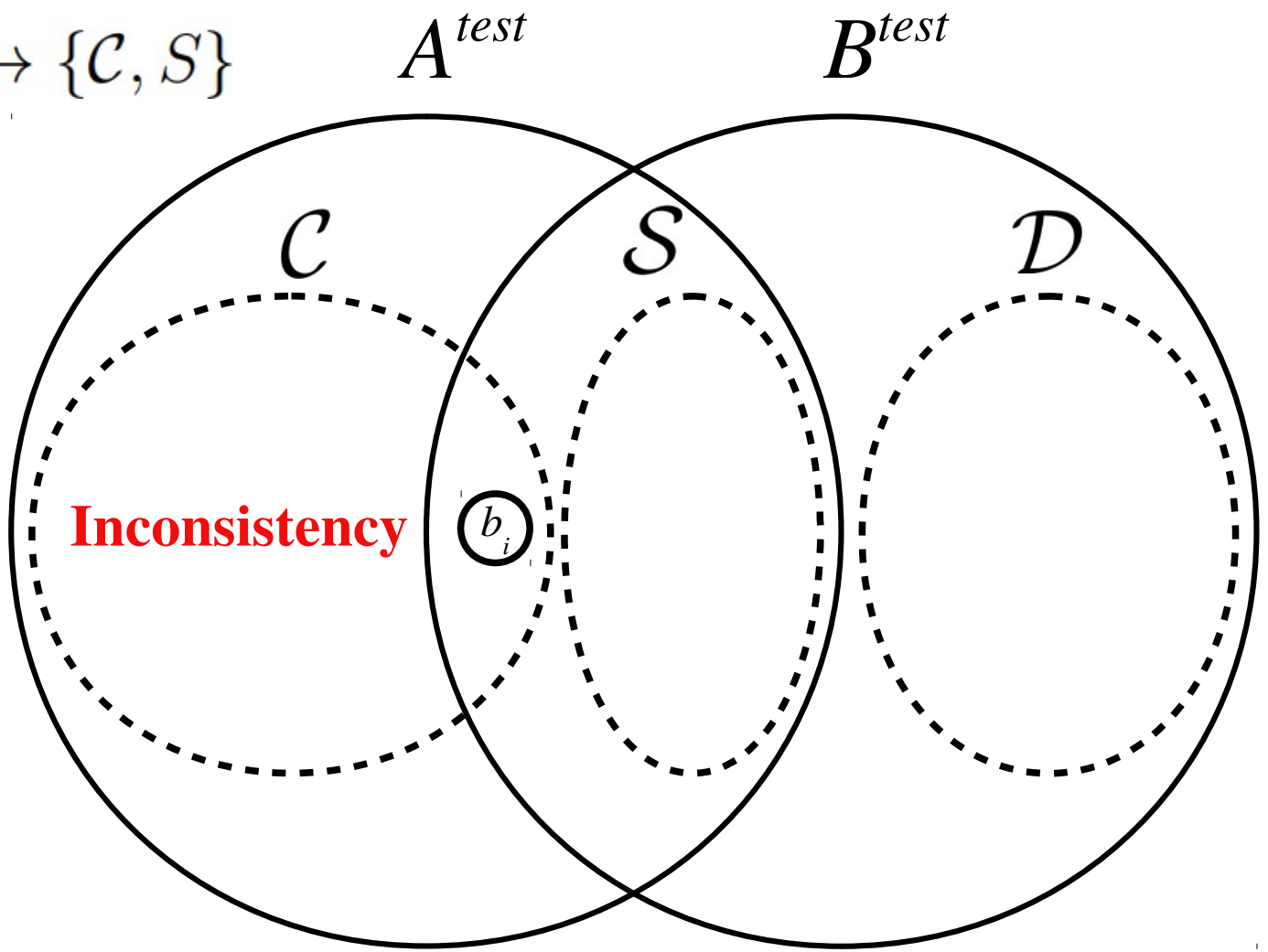


F2 Inconsistencies

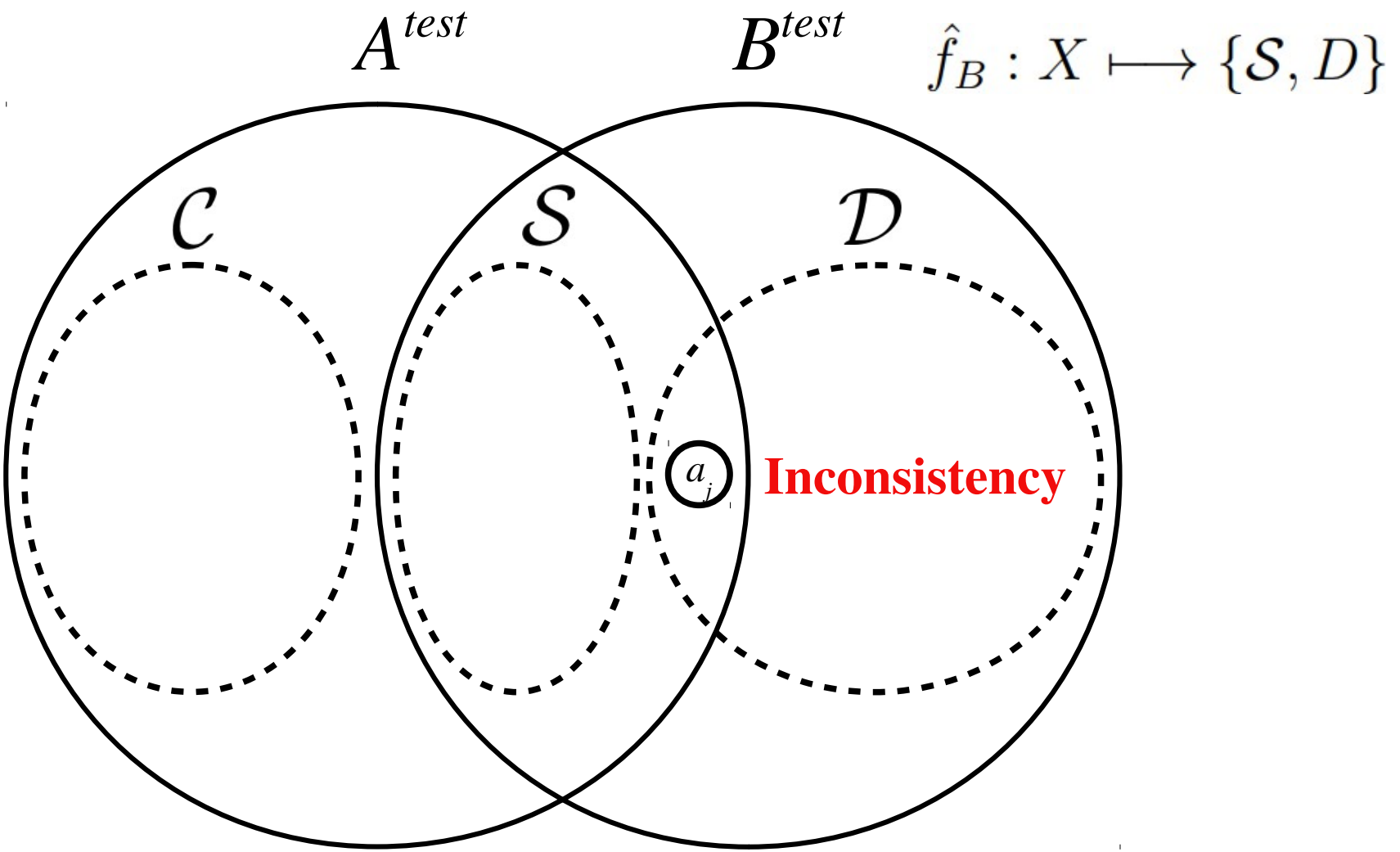
$$F_2(i) \equiv \begin{cases} \text{If } \hat{f}_B(a_i) \neq \mathcal{S}_B, & \text{then output “inconsistency”,} \\ \text{If } \hat{f}_A(b_i) \neq \mathcal{S}_A, & \text{then output “inconsistency”.} \end{cases}$$

F2 Inconsistencies

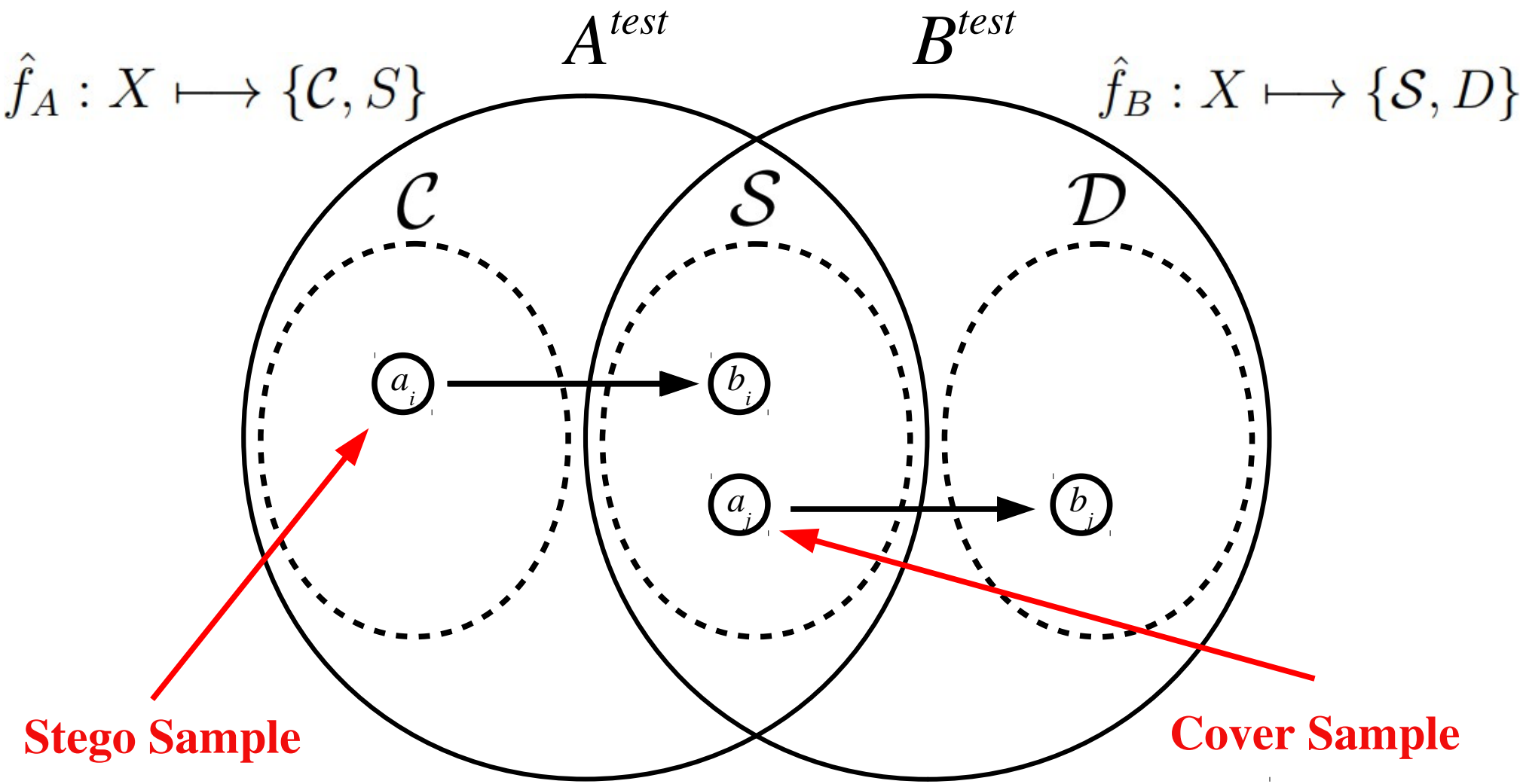
$\hat{f}_A : X \longmapsto \{C, S\}$



F2 Inconsistencies



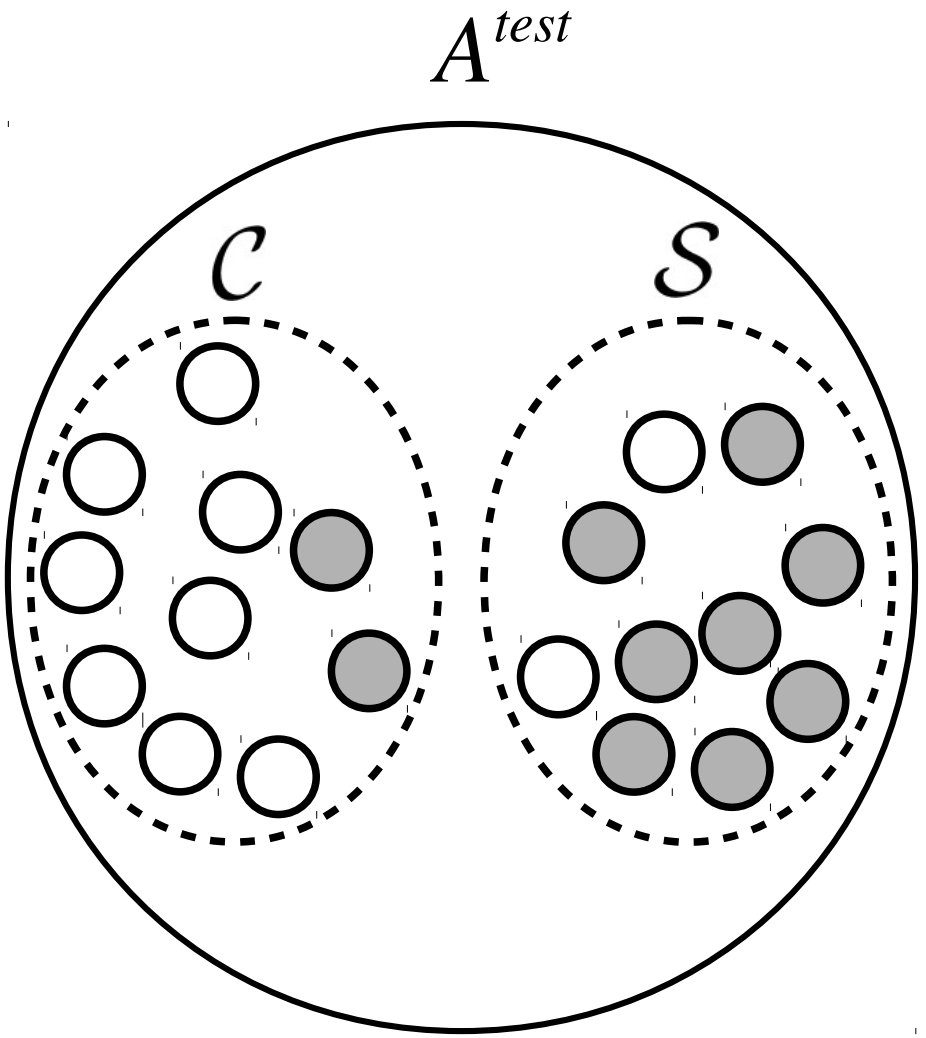
Undetectable Inconsistencies



Outline

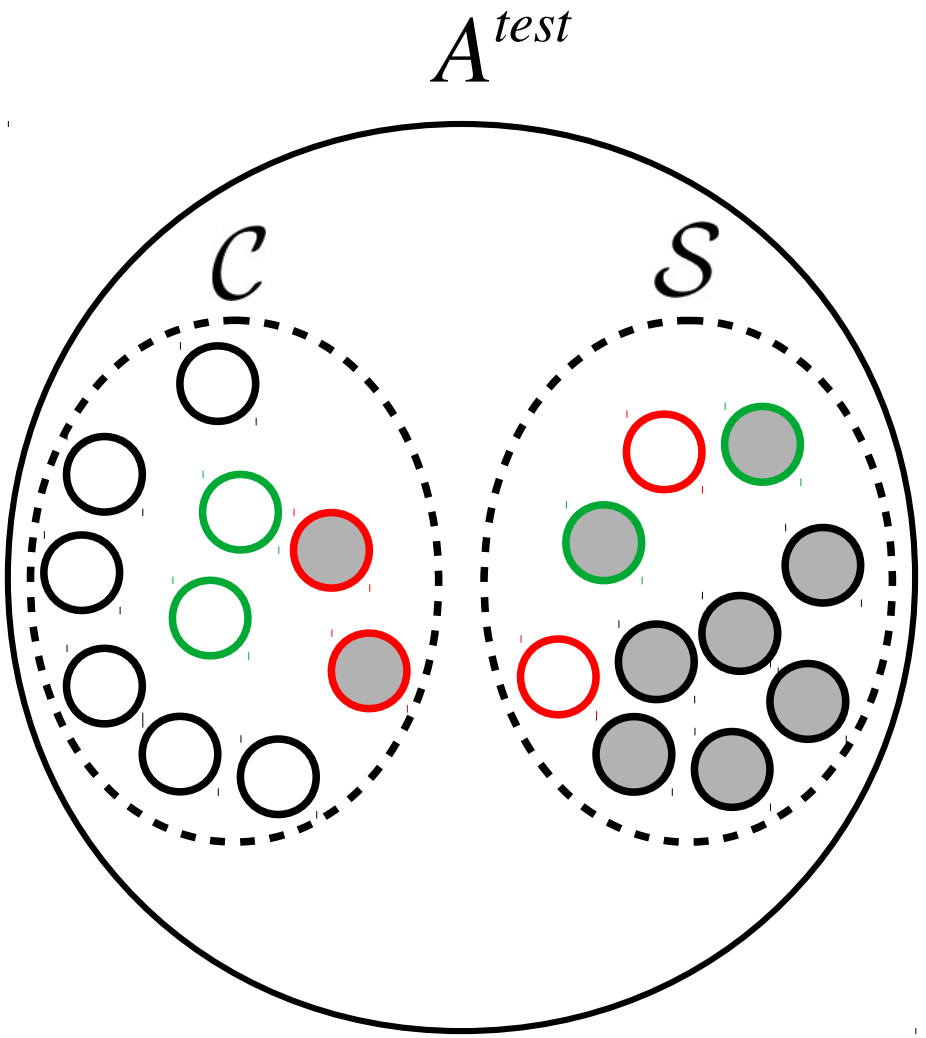
1. Introduction
2. Training and testing sets
3. Detection of inconsistencies
4. **Prediction of the classifier's error**
5. Experimental results
6. Conclusions and future work

Prediction of the Classifier's Error



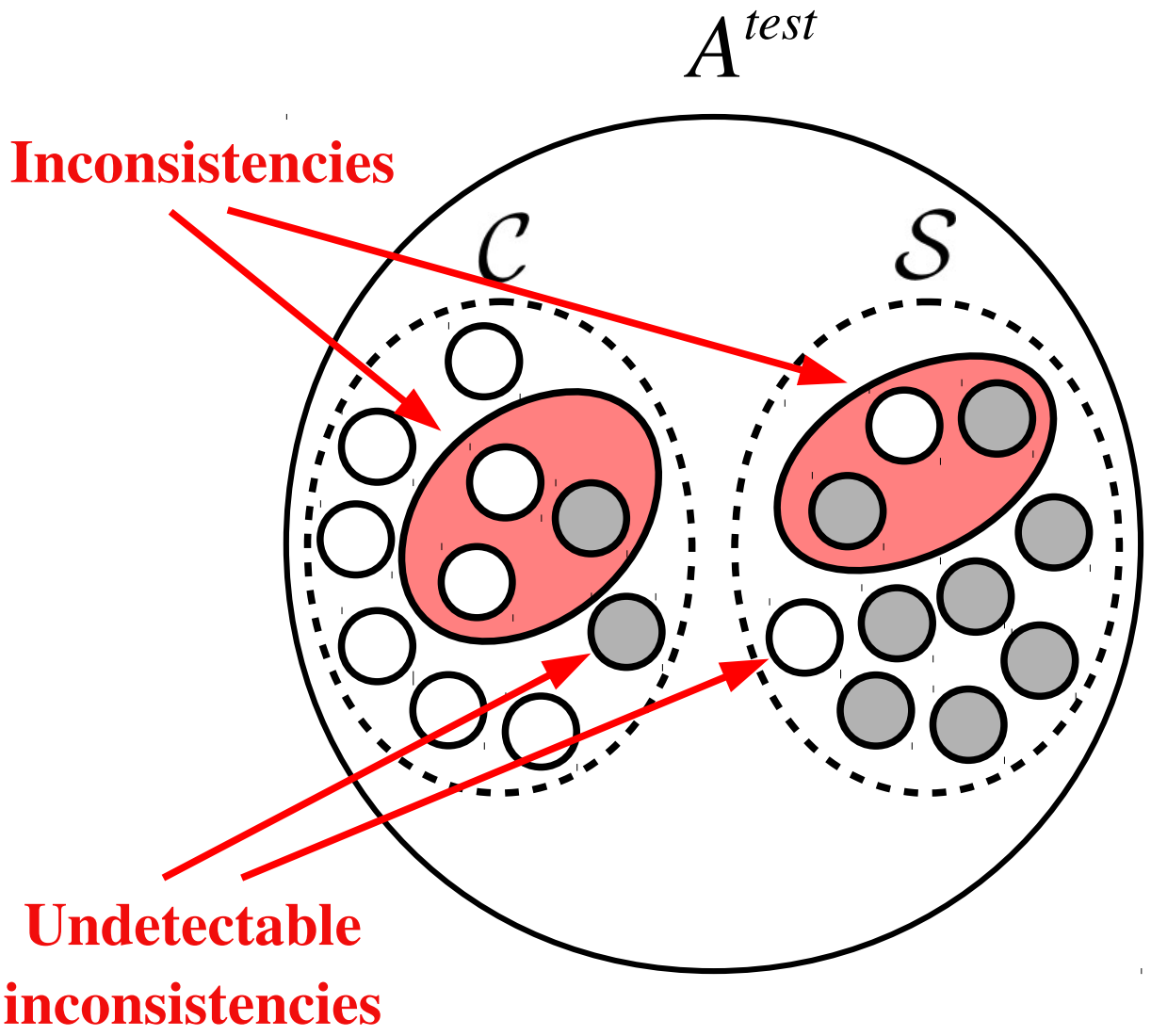
Err = 4/20 = 0.2

Prediction of the Classifier's Error



Err = 4/20 = 0.2

Prediction of the Classifier's Error



Err = 4/20 = 0.2

$$\mathrm{Err}_{\mathrm{pred}} = \frac{\mathrm{INC}}{2\,|A^{\mathrm{test}}|}$$

$\mathrm{Err}_{\mathrm{pred}}$ = 6 / (2·20) = 0.15

Outline

1. Introduction
2. Training and testing sets
3. Detection of inconsistencies
4. Prediction of the classifier's error
5. **Experimental results**
6. Conclusions and future work

Experiments

ALGO	DBs	C/S	CLF	Err	INC	Err_{pred}
HILL-0.40	BOSS/BOSS	500/500	SRNET	**0.2520**	527	**0.2635**
HILL-0.40	BOSS/BOWS2	500/500	SRNET	**0.2600**	571	**0.2855**
HILL-0.40	BOSS/ALASKA	500/500	SRNET	**0.3840**	765	**0.3825**
HILL-0.40	BOWS2/BOWS2	500/500	SRNET	**0.2670**	544	**0.2720**
HILL-0.40	BOWS2/BOSS	500/500	SRNET	**0.3570**	650	**0.3250**
HILL-0.40	BOWS2/ALASKA	500/500	SRNET	**0.3880**	761	**0.3805**
HILL-0.40	ALASKA/ALASKA	500/500	SRNET	**0.3940**	782	**0.3910**
HILL-0.40	ALASKA/BOWS2	500/500	SRNET	**0.3930**	786	**0.3930**
HILL-0.40	ALASKA/BOSS	500/500	SRNET	**0.3900**	824	**0.4120**

Experiments

$$\mathrm{Err}_{\mathrm{pred}} = \frac{\mathrm{INC}}{2\,|A^{\mathrm{test}}|}$$

ALGO	DBs	C/S	CLF	Err	INC	Err_{pred}
HILL-0.40	BOSS/BOSS	500/500	SRNET	**0.2520**	527	**0.2635**
HILL-0.40	BOSS/BOWS2	500/500	SRNET	**0.2600**	571	**0.2855**
HILL-0.40	BOSS/ALASKA	500/500	SRNET	**0.3840**	765	**0.3825**
HILL-0.40	BOWS2/BOWS2	500/500	SRNET	**0.2670**	544	**0.2720**
HILL-0.40	BOWS2/BOSS	500/500	SRNET	**0.3570**	650	**0.3250**
HILL-0.40	BOWS2/ALASKA	500/500	SRNET	**0.3880**	761	**0.3805**
HILL-0.40	ALASKA/ALASKA	500/500	SRNET	**0.3940**	782	**0.3910**
HILL-0.40	ALASKA/BOWS2	500/500	SRNET	**0.3930**	786	**0.3930**
HILL-0.40	ALASKA/BOSS	500/500	SRNET	**0.3900**	824	**0.4120**

Experiments

ALGO	DBs	C/S	CLF	Err	INC	Err_{pred}
HILL-0.40	BOSS/BOSS	500/500	SRNET	**0.2520**	527	**0.2635**
HILL-0.40	BOSS/BOWS2	500/500	SRNET	**0.2600**	571	**0.2855**
HILL-0.40	BOSS/ALASKA	500/500	SRNET	**0.3840**	765	**0.3825**
HILL-0.40	BOWS2/BOWS2	500/500	SRNET	**0.2670**	544	**0.2720**
HILL-0.40	BOWS2/BOSS	500/500	SRNET	**0.3570**	650	**0.3250**
HILL-0.40	BOWS2/ALASKA	500/500	SRNET	**0.3880**	761	**0.3805**
HILL-0.40	ALASKA/ALASKA	500/500	SRNET	**0.3940**	782	**0.3910**
HILL-0.40	ALASKA/BOWS2	500/500	SRNET	**0.3930**	786	**0.3930**
HILL-0.40	ALASKA/BOSS	500/500	SRNET	**0.3900**	824	**0.4120**

Experiments

ALGO	DBs	C/S	CLF	Err	Err_{pred}
UED-0.40	BOSS-J95/BOSS-J95	500/500	GFR+EC	**0.1530**	**0.1310**
UED-0.40	BOSS-J95/BOWS2-J95	500/500	GFR+EC	**0.1900**	**0.1635**
UED-0.40	BOSS-J95/ALASKA-J95	500/500	GFR+EC	**0.4310**	**0.4145**
J-UNIW-0.40	BOSS-J95/BOSS-J95	500/500	GFR+EC	**0.2280**	**0.2295**
J-UNIW-0.40	BOSS-J95/BOWS2-J95	500/500	GFR+EC	**0.2640**	**0.2560**
UED-0.40	BOWS2-J95/BOWS2-J95	500/500	GFR+EC	**0.1660**	**0.1525**
UED-0.40	BOWS2-J95/BOSS-J95	500/500	GFR+EC	**0.1690**	**0.1460**
UED-0.40	BOWS2-J95/ALASKA-J95	500/500	GFR+EC	**0.4180**	**0.3995**
J-UNIW-0.40	BOWS2-J95/BOWS2-J95	500/500	GFR+EC	**0.2600**	**0.2380**
J-UNIW-0.40	BOWS2-J95/BOSS-J95	500/500	GFR+EC	**0.2460**	**0.2380**
UED-0.40	ALASKA-J95/ALASKA-J95	500/500	GFR+EC	**0.3040**	**0.2665**
UED-0.40	ALASKA-J95/BOSS-J95	500/500	GFR+EC	**0.2350**	**0.2065**
UED-0.40	ALASKA-J95/BOWS2-J95	500/500	GFR+EC	**0.2400**	**0.2100**

Experiments

ALGO	DBs	C/S	CLF	Err	Err_{pred}
HILL-0.40	BOSS/BOSS	500/500	RM+EC	**0.2440**	**0.2410**
HILL-0.40	BOSS/BOSS	500/250	RM+EC	**0.2573**	**0.2407**
HILL-0.40	BOSS/BOSS	500/0	RM+EC	**0.2840**	**0.2360**
HILL-0.40	BOSS/BOSS	250/500	RM+EC	**0.2320**	**0.2420**
HILL-0.40	BOSS/BOSS	0/500	RM+EC	**0.2040**	**0.2460**
HILL-0.40	BOSS/BOWS2	500/500	RM+EC	**0.4530**	**0.4365**
HILL-0.40	BOSS/BOWS2	500/250	RM+EC	**0.6027**	**0.4233**
HILL-0.40	BOSS/BOWS2	500/0	RM+EC	**0.8920**	**0.3950**
HILL-0.40	BOSS/BOWS2	250/500	RM+EC	**0.3067**	**0.4473**
HILL-0.40	BOSS/BOWS2	0/500	RM+EC	**0.0140**	**0.4780**
HILL-0.40	BOSS/ALASKA	500/500	RM+EC	**0.4810**	**0.4750**
HILL-0.40	BOSS/ALASKA	500/250	RM+EC	**0.5453**	**0.4727**
HILL-0.40	BOSS/ALASKA	500/0	RM+EC	**0.7000**	**0.4710**
HILL-0.40	BOSS/ALASKA	250/500	RM+EC	**0.4027**	**0.4787**
HILL-0.40	BOSS/ALASKA	0/500	RM+EC	**0.2620**	**0.4790**

Experiments

ALGO	DBs	C/S	CLF	Err	Err_{pred}
HILL-0.40	BOSS/BOSS	500/500	RM+EC	**0.2440**	**0.2410**
HILL-0.40	BOSS/BOSS	500/250	RM+EC	**0.2573**	**0.2407**
HILL-0.40	BOSS/BOSS	500/0	RM+EC	**0.2840**	**0.2360**
HILL-0.40	BOSS/BOSS	250/500	RM+EC	**0.2320**	**0.2420**
HILL-0.40	BOSS/BOSS	0/500	RM+EC	**0.2040**	**0.2460**
HILL-0.40	BOSS/BOWS2	500/500	RM+EC	**0.4530**	**0.4365**
HILL-0.40	BOSS/BOWS2	500/250	RM+EC	**0.6027**	**0.4233**
HILL-0.40	BOSS/BOWS2	500/0	RM+EC	**0.8920**	**0.3950**
HILL-0.40	BOSS/BOWS2	250/500	RM+EC	**0.3067**	**0.4473**
HILL-0.40	BOSS/BOWS2	0/500	RM+EC	**0.0140**	**0.4780**
HILL-0.40	BOSS/ALASKA	500/500	RM+EC	**0.4810**	**0.4750**
HILL-0.40	BOSS/ALASKA	500/250	RM+EC	**0.5453**	**0.4727**
HILL-0.40	BOSS/ALASKA	500/0	RM+EC	**0.7000**	**0.4710**
HILL-0.40	BOSS/ALASKA	250/500	RM+EC	**0.4027**	**0.4787**
HILL-0.40	BOSS/ALASKA	0/500	RM+EC	**0.2620**	**0.4790**

Experiments

ALGO	DBs	C/S	CLF	Err	Err_{pred}
HILL-0.40	BOSS/BOSS	500/500	RM+EC	**0.2440**	**0.2410**
HILL-0.40	BOSS/BOSS	500/250	RM+EC	**0.2573**	**0.2407**
HILL-0.40	BOSS/BOSS	500/0	RM+EC	**0.2840**	**0.2360**
HILL-0.40	BOSS/BOSS	250/500	RM+EC	**0.2320**	**0.2420**
HILL-0.40	BOSS/BOSS	0/500	RM+EC	**0.2040**	**0.2460**
HILL-0.40	BOSS/BOWS2	500/500	RM+EC	**0.4530**	**0.4365**
HILL-0.40	BOSS/BOWS2	500/250	RM+EC	**0.6027**	**0.4233**
HILL-0.40	BOSS/BOWS2	500/0	RM+EC	**0.8920**	**0.3950**
HILL-0.40	BOSS/BOWS2	250/500	RM+EC	**0.3067**	**0.4473**
HILL-0.40	BOSS/BOWS2	0/500	RM+EC	**0.0140**	**0.4780**
HILL-0.40	BOSS/ALASKA	500/500	RM+EC	**0.4810**	**0.4750**
HILL-0.40	BOSS/ALASKA	500/250	RM+EC	**0.5453**	**0.4727**
HILL-0.40	BOSS/ALASKA	500/0	RM+EC	**0.7000**	**0.4710**
HILL-0.40	BOSS/ALASKA	250/500	RM+EC	**0.4027**	**0.4787**
HILL-0.40	BOSS/ALASKA	0/500	RM+EC	**0.2620**	**0.4790**

Experiments

ALGO	DBs	C/S	CLF	Err	Err_{pred}
HILL-0.40	BOSS/BOSS	500/500	RM+EC	**0.2440**	**0.2410**
HILL-0.40	BOSS/BOSS	500/250	RM+EC	**0.2573**	**0.2407**
HILL-0.40	BOSS/BOSS	500/0	RM+EC	**0.2840**	**0.2360**
HILL-0.40	BOSS/BOSS	250/500	RM+EC	**0.2320**	**0.2420**
HILL-0.40	BOSS/BOSS	0/500	RM+EC	**0.2040**	**0.2460**
HILL-0.40	BOSS/BOWS2	500/500	RM+EC	**0.4530**	**0.4365**
HILL-0.40	BOSS/BOWS2	500/250	RM+EC	**0.6027**	**0.4233**
HILL-0.40	BOSS/BOWS2	500/0	RM+EC	**0.8920**	**0.3950**
HILL-0.40	BOSS/BOWS2	250/500	RM+EC	**0.3067**	**0.4473**
HILL-0.40	BOSS/BOWS2	0/500	RM+EC	**0.0140**	**0.4780**
HILL-0.40	BOSS/ALASKA	500/500	RM+EC	**0.4810**	**0.4750**
HILL-0.40	BOSS/ALASKA	500/250	RM+EC	**0.5453**	**0.4727**
HILL-0.40	BOSS/ALASKA	500/0	RM+EC	**0.7000**	**0.4710**
HILL-0.40	BOSS/ALASKA	250/500	RM+EC	**0.4027**	**0.4787**
HILL-0.40	BOSS/ALASKA	0/500	RM+EC	**0.2620**	**0.4790**

Outline

1. Introduction
2. Training and testing sets
3. Detection of inconsistencies
4. Prediction of the classifier's error
5. Experimental results
6. **Conclusions and future work**

Conclusions

SUMMARY

- Batch Steganography & Pooled Steganalysis
- Attack to known algorithm and bit rate
- Prediction of the classifier's error

FUTURE WORK

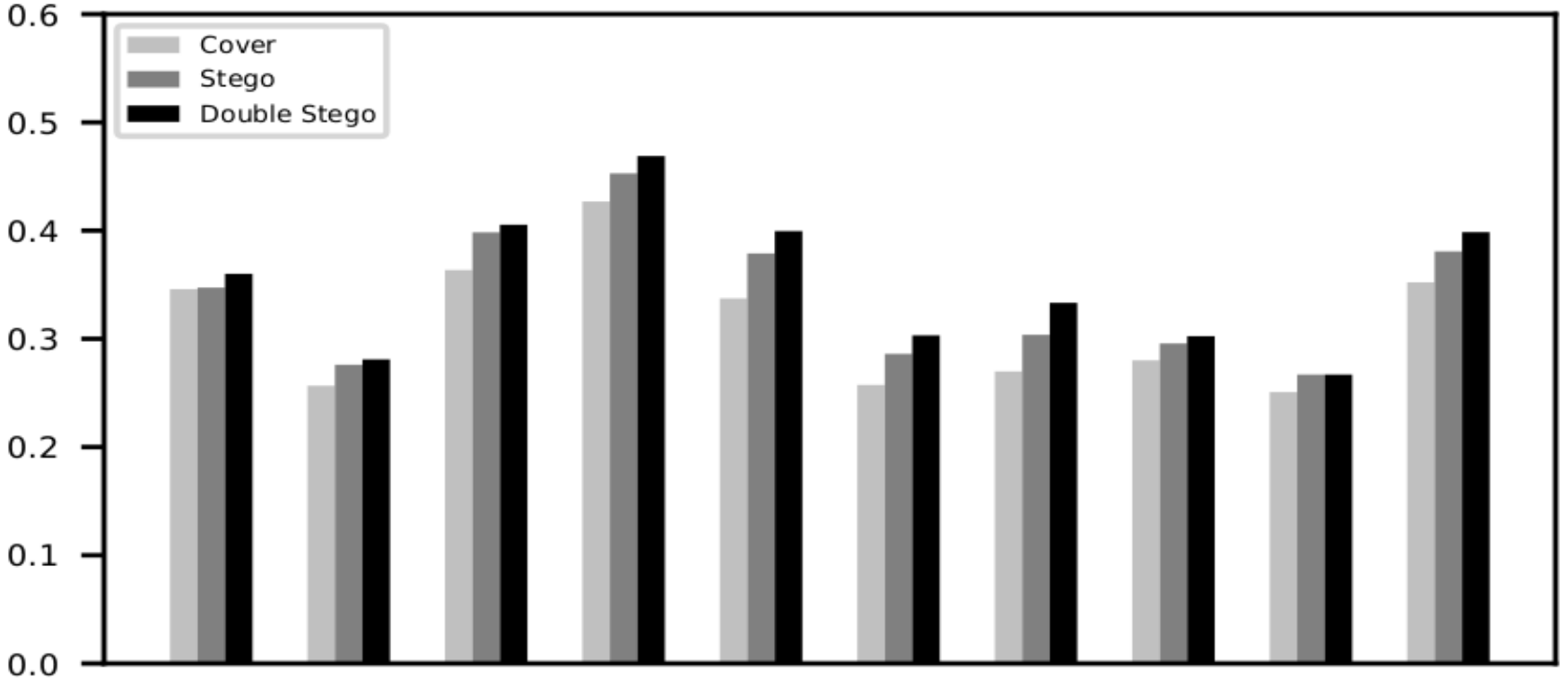
- Stego Source Mismatch
- Single images

Unbalanced Prediction

$$p = \frac{\#C}{|A_{test}|}, \quad p = \frac{\#S}{|A_{test}|} = 1 - p, \quad ERR_{pred} = \frac{q \cdot INC_C + p \cdot INC_S}{|A_{test}|}$$

ALGO	DBs	C/S	CLF	Err	Err_{pred}	INC	INC_C	INC_S	
HILL-0.40	BOSS/BOWS2	500/500	RM+EC	**0.4530**	**0.4365**	873	44	829	**0.4365**
HILL-0.40	BOSS/BOWS2	500/250	RM+EC	**0.6027**	**0.4233**	635	43	592	**0.5453**
HILL-0.40	BOSS/BOWS2	500/0	RM+EC	**0.8920**	**0.3950**	395	38	357	**0.7140**
HILL-0.40	BOSS/BOWS2	250/500	RM+EC	**0.3067**	**0.4473**	671	24	647	**0.3088**
HILL-0.40	BOSS/BOWS2	0/500	RM+EC	**0.0140**	**0.4780**	478	6	472	**0.0120**

Feature embedding



Top 10 features

HILL 0.4 / Bossbase

Stego / Double Stego

ALGO	BR	1 embedding ±1	1 embedding ±2	2 embeddings ±1	2 embeddings ±2
HILL	0.4	22,582,706	0	33,993,485	2,819,705
HILL	0.2	9,897,485	0	16,112,459	933,376
UNIWARD	0.4	19,509,940	0	32,748,639	1,563,635
UNIWARD	0.2	8,523,446	0	15,139,023	477,200
LSBM	0.2	27,528,954	0	49,277,814	1,439,498

1000 Bossbase images

Experiment HILL 0.40:
B set with +/- 1 and +/-2, Error: 0.2770
B set with +/- 1 and +/-1, Error: 0.3160

Error (consistent samples)

ALGO	DBs	C/S	CLF	Err	Err_{pred}	Err	TP	TN	FP	FN	INC
HILL-0.40	BOSS/BOSS	500/500	RM+EC	**0.2440**	**0.2410**	0.1564	214	223	41	40	482
HILL-0.40	BOSS/BOSS	500/250	RM+EC	**0.2573**	**0.2407**	0.1491	108	223	41	17	361
HILL-0.40	BOSS/BOSS	500/0	RM+EC	**0.2840**	**0.2360**	0.1553	0	223	41	0	236
HILL-0.40	BOSS/BOSS	250/500	RM+EC	**0.2320**	**0.2420**	0.1628	214	110	23	40	363
HILL-0.40	BOSS/BOSS	0/500	RM+EC	**0.2040**	**0.2460**	0.1575	214	0	0	40	246
HILL-0.40	BOSS/BOWS2	500/500	RM+EC	**0.4530**	**0.4365**	0.7087	21	16	89	1	873
HILL-0.40	BOSS/BOWS2	500/250	RM+EC	**0.6027**	**0.4233**	0.7826	9	16	89	1	635
HILL-0.40	BOSS/BOWS2	500/0	RM+EC	**0.8920**	**0.3950**	0.8476	0	16	89	0	395
HILL-0.40	BOSS/BOWS2	250/500	RM+EC	**0.3067**	**0.4473**	0.6203	21	9	48	1	671
HILL-0.40	BOSS/BOWS2	0/500	RM+EC	**0.0140**	**0.4780**	0.0455	21	0	0	1	478
HILL-0.40	BOSS/ALASKA	500/500	RM+EC	**0.4810**	**0.4750**	0.2600	19	18	11	2	950
HILL-0.40	BOSS/ALASKA	500/250	RM+EC	**0.5453**	**0.4727**	0.2927	11	18	11	1	709
HILL-0.40	BOSS/ALASKA	500/0	RM+EC	**0.7000**	**0.4710**	0.3793	0	18	11	0	471
HILL-0.40	BOSS/ALASKA	250/500	RM+EC	**0.4027**	**0.4787**	0.1875	19	7	4	2	718
HILL-0.40	BOSS/ALASKA	0/500	RM+EC	**0.2620**	**0.4790**	0.0952	19	0	0	2	479

Stego Source Mismatch

ALGO	DBs	C/S	CLF	Err	Err_{pred}
HILL-0.40/0.20	BOSS/BOSS	500/500	RM+EC	**0.2850**	**0.4200**
HILL-0.40/0.30	BOSS/BOSS	500/500	RM+EC	**0.2450**	**0.3470**
HILL-0.40/0.40	BOSS/BOSS	500/500	RM+EC	**0.2440**	**0.2410**
HILL-0.40/0.50	BOSS/BOSS	500/500	RM+EC	**0.2640**	**0.1880**
HILL-0.40/0.60	BOSS/BOSS	500/500	RM+EC	**0.2820**	**0.1800**