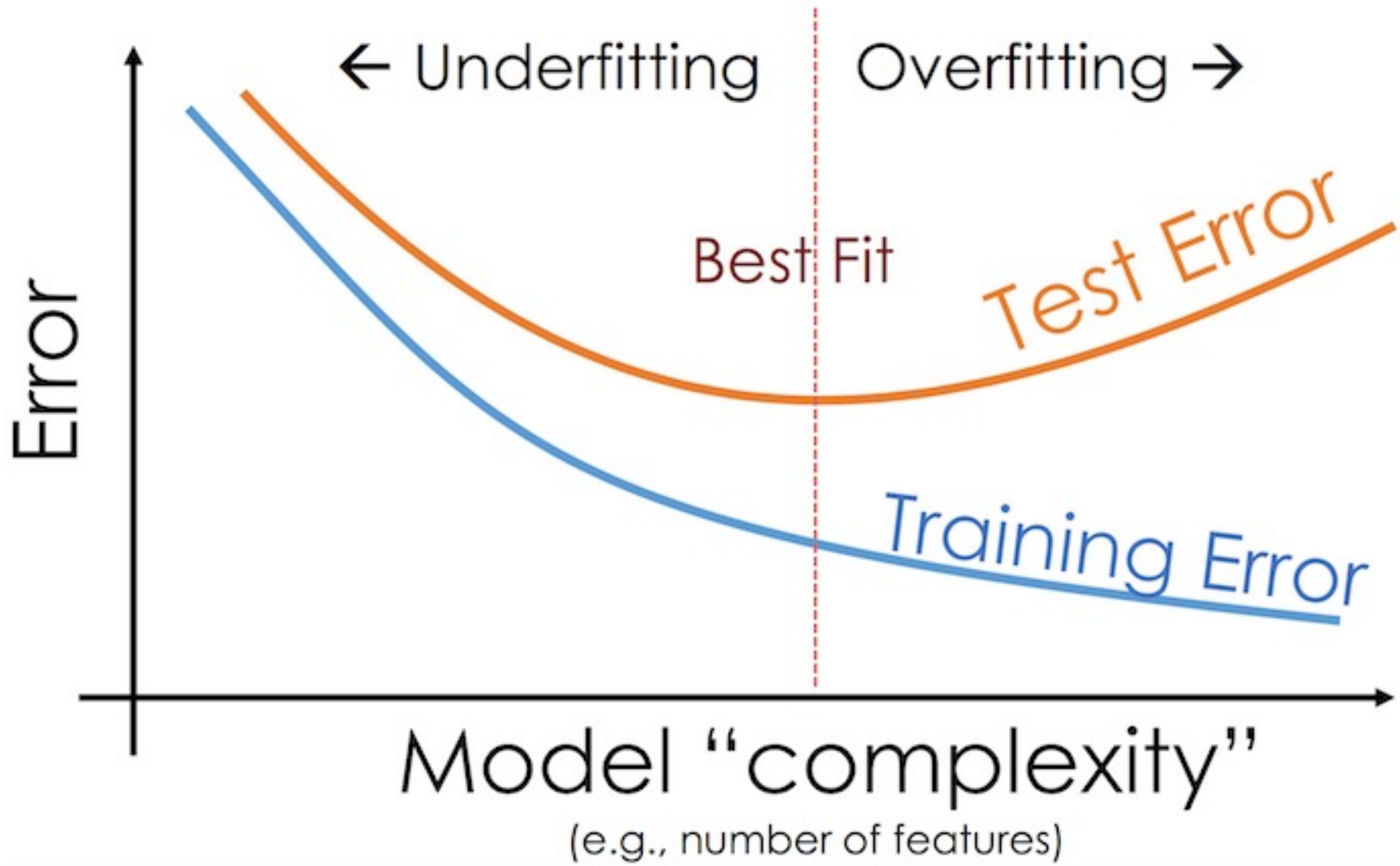
Model Verification

Exercise 7

Model Verification

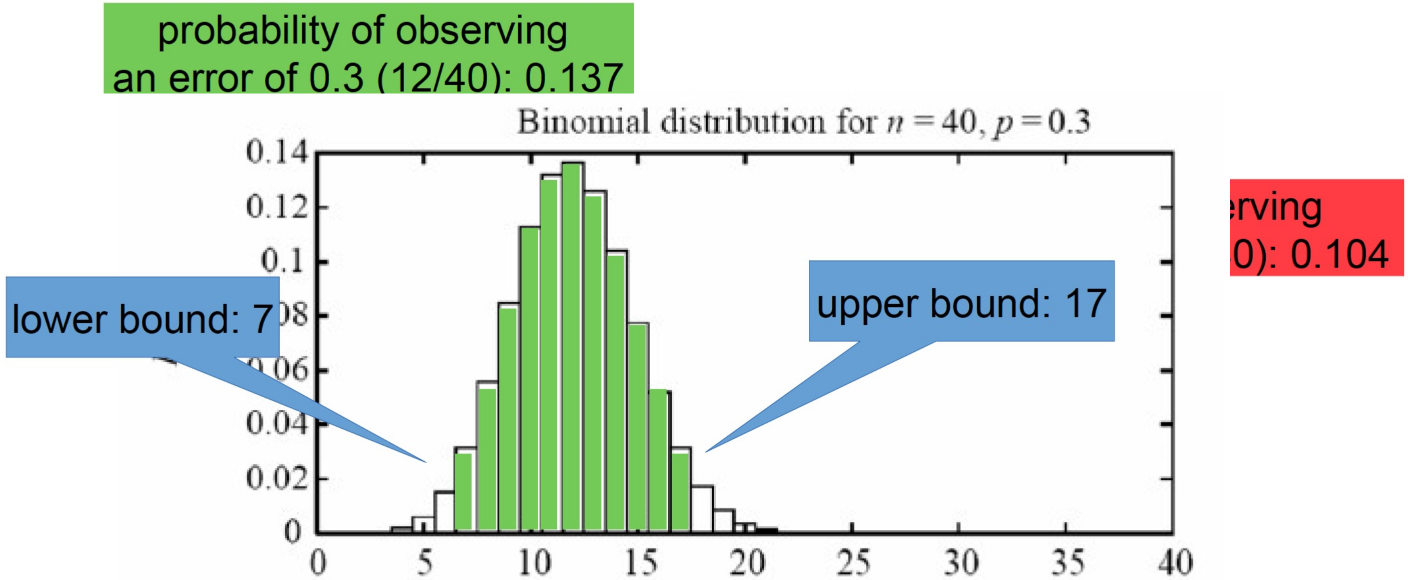
Training vs. Test Error



https://www.textbook.ds100.org/ch/20/bias_cv.html

Model Verification
Confidence Intervals

Caution: only for sample size > 30!



With p% probability, $error_D$ is in [$error_S$ – y, $error_S$ + y]

With y= $z_N \cdot \sqrt{\frac{error_S(1-error_S)}{n}}$

$N\%$:	50%	68%	80%	90%	95%	98%	99%
z_N:	0.67	1.00	1.28	1.64	1.96	2.33	2.58

→ With 95% probability, $error_D$ is in
[0.3 – 0.142 ; 0.3 + 0.142]
= [0.158 ; 0.442]

Model Verification

Confidence Intervals

$N\%$:	50%	68%	80%	90%	95%	98%	99%
z_N:	0.67	1.00	1.28	1.64	1.96	2.33	2.58

TASK

You are using a machine learning solution from the company Flancrest Enterprises. Recently, you were contacted by the Junior Vice President of CompuGlobalHyperMegaNet and he offered you to switch to his solution. As a migration is very costly, you only want to switch if you can be at least 90% sure that the new solution is better. For such purposes, you have a dedicated test set with 420 examples where your current solution makes 105 errors. What is the highest number of errors that you accept for the new solution in order to switch?

Model Verification
Confidence Intervals

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

$N\%$:	50%	68%	80%	90%	95%	98%	99%
z_N:	0.67	1.00	1.28	1.64	1.96	2.33	2.58

|S| = 420 (> 30, so we can use z-test!)

$$y = z_N \cdot \sqrt{\frac{error_S(1-error_S)}{n}}$$

$error_S = 0.25$

$z_N = 1.64$

$$y = 1.64 * \sqrt{\frac{0.25 * (1 - 0.25)}{420}} = 1.64 * 0.02 = 0.0328$$

→ With 90% probability, $error_D$ is in [0.2172 ; 0.2828]

→ The maxmimum number of errors for the new solution is $\lfloor 0.2172 * 420 \rfloor = 91$.

Sign Test

- Methods M (new) and S (SotA)
- Count wins, losses, and ties of M with respect to S
- Given significance level alpha (typically 0.05 or 0.1), check how many wins M needs to be significantly better
- For *n* = 9 and *alpha* = 0.05:
  If M has at least 8 wins, it is signficantly better

#data sets	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$w_{0.05}$	5	6	7	7	8	9	9	10	10	11	12	12	13	13	14	15	15	16	17	18	18
$w_{0.10}$	5	6	6	7	7	8	9	9	10	10	11	12	12	13	13	14	14	15	16	16	17

Wilcoxon Signed-Rank Test

- Methods M (new) and S (SotA)
- Compute deltas of methods
- Sort examples by absolute delta
- Assign ranks to examples (highest rank for highest delta)
- R+ is the sum of ranks won by M
- R- is the sum of ranks won by S
- For *n* = 12 and *alpha* = 0.05, M is significantly better than S if R- < 17

n	$\alpha_{\text{two-tailed}} \leq 0.10$ $\alpha_{\text{one-tailed}} \leq 0.05$	$\alpha_{\text{two-tailed}} \leq 0.05$ $\alpha_{\text{one-tailed}} \leq 0.025$	$\alpha_{\text{two-tailed}} \leq 0.02$ $\alpha_{\text{one-tailed}} \leq 0.01$	$\alpha_{\text{two-tailed}} \leq 0.01$ $\alpha_{\text{one-tailed}} \leq 0.005$
5	0			
6	2	0		
7	3	2	0	
8	5	3	1	0
9	8	5	3	1
10	10	8	5	3
11	13	10	7	5
12	17	13	9	7
13	21	17	12	9
14	25	21	15	12
15	30	25	19	15
16	35	29	23	19
17	41	34	27	23
18	47	40	32	27
19	53	46	37	32
20	60	52	43	37
21	67	58	49	42
22	75	65	55	48
23	83	73	62	54
24	91	81	69	61
25	100	89	76	68
26	110	98	84	75
27	119	107	92	83
28	130	116	101	91
29	140	126	110	100
30	151	137	120	109

Source: Adapted from McComack, R. L. (1965). Extended tables of the Wilcoxon matched pair signed rank statistic. *Journal of the American Statistical Association*, *60*, 864–871. Reprinted with permission from *The Journal of the American Statistical Association*.

Model Verification

Sign test & Wilcoxon signed-rank test

TASK

Determine whether the new variant is significantly better than the old variant at a significance level of alpha = 0.05

a) Using a sign test

b) Using a Wilcoxon signed-rank test

Problem	New	Old
1	0.83	0.73
2	0.67	0.72
3	0.29	0.27
4	0.47	0.41
5	0.57	0.43
6	0.35	0.22
7	0.47	0.36
8	0.57	0.53
9	0.89	0.89
10	0.22	0.31
11	0.57	0.54
12	0.15	0.12
13	0.39	0.46
14	0.23	0.21
avg.	0.48	0.44

Model Verification

Sign test & Wilcoxon signed-rank test

Problem	New	Old	Delta	Delta (abs.)	Rank	R+	R-

n	Two-Tailed Test		One-Tailed Test	
	α = .05	α = .01	α = .05	α = .01
5	--	--	0	--
6	0	--	2	--
7	2	--	3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15
15	25	15	30	19
16	29	19	35	23
17	34	23	41	27
18	40	27	47	32
19	46	32	53	37
20	52	37	60	43

Sign test

10 wins, 3 losses, 1 tie

#datasets = 13, $w_{0.05}$ = 10

→ result **is significant**

#data sets	10	11	12	13	14	15	16
$w_{0.05}$	9	9	10	10	11	12	12
$w_{0.10}$	8	9	9	10	10	11	12

#datasets = 13

$\alpha_{\text{one-tailed}} \leqq 0.05$ = 21

→ R- is not smaller than 21

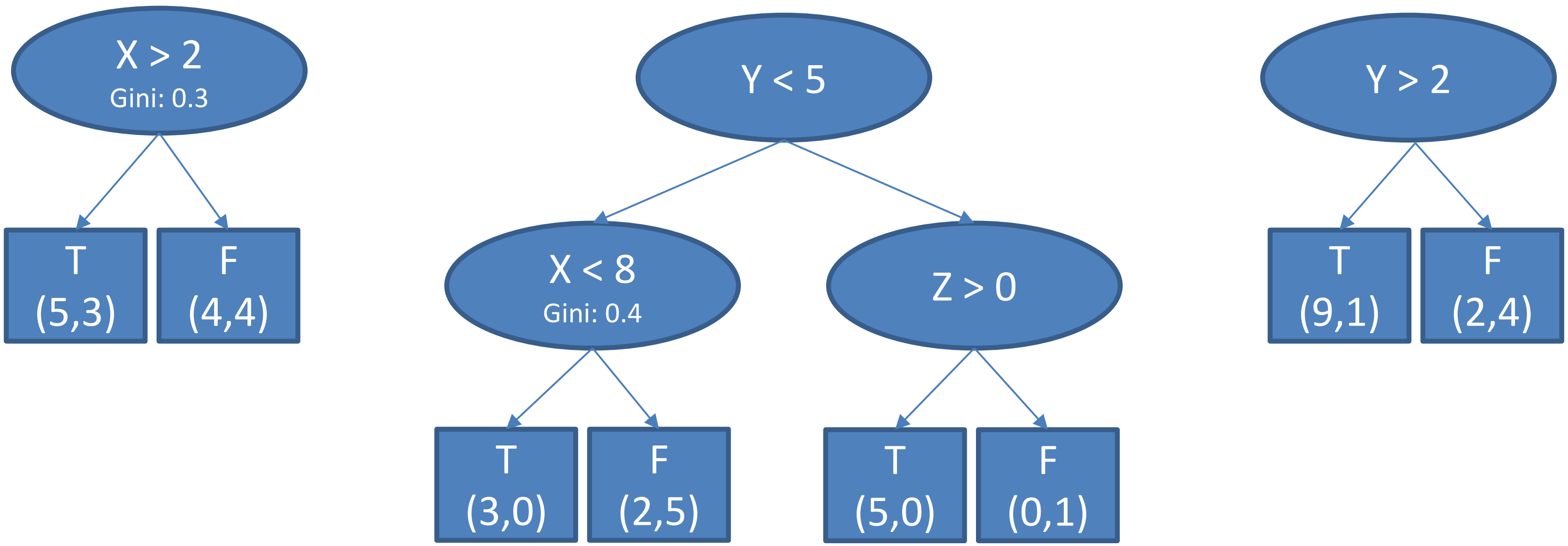
→ result **is not significant**

Measuring Feature Importance

TASK

Compute the importance of feature X given a Random Forest Classifier consisting of the following trees:

For convenience, use the provided Gini-Index values.



Measuring Feature Importance

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

$$\text{Importance}(X) = \frac{(1 * 0.3) + (\frac{5}{8} * 0.4) + 0}{3} = 0.1833$$

$p(n) = 1$

$\Delta I(s_n,n) = 0.3$

$p(n) = \frac{5}{8}$

$\Delta I(s_n,n) = 0.4$

$\cancel{p(n) =}$

$\cancel{\Delta I(s_n,n) =}$