Minimizing Average Cost per Patient for Thyroid Gland Disorder Diagnosis

By Vince Wu and Ryan Liu

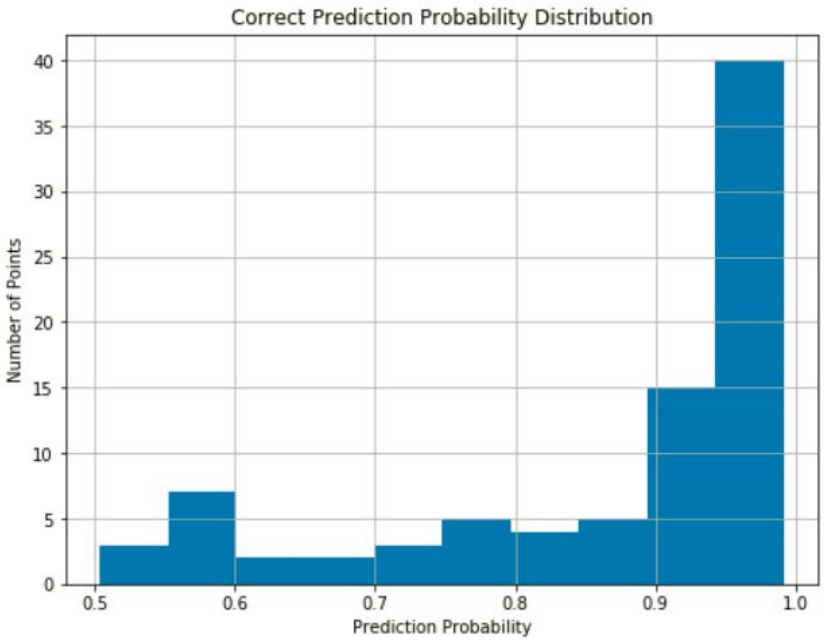
COSTS FUNCTION

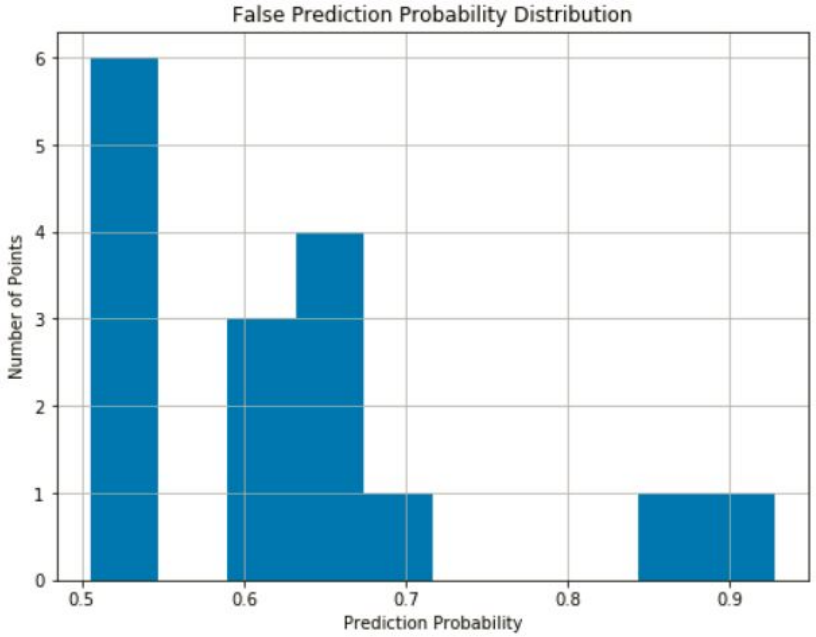
- Correct Diagnosis (=C) - $0
- Misdiagnosis (=M) - $5000
- Abstain Prediction (=A) (forward to endocrinologist) - $1000
- Total Number of Patients (N = C+M+A)

- Total Costs:

$$T = \frac{5000M + 1000A}{N}$$

PROBABILITY ANALYSIS





- OvR (by default threshold) assign classification with $p > 0.5$
- Larger p → Higher chance for correct prediction (diagnosis)
- Smaller p → More likely for misdiagnosis

NEW MODEL

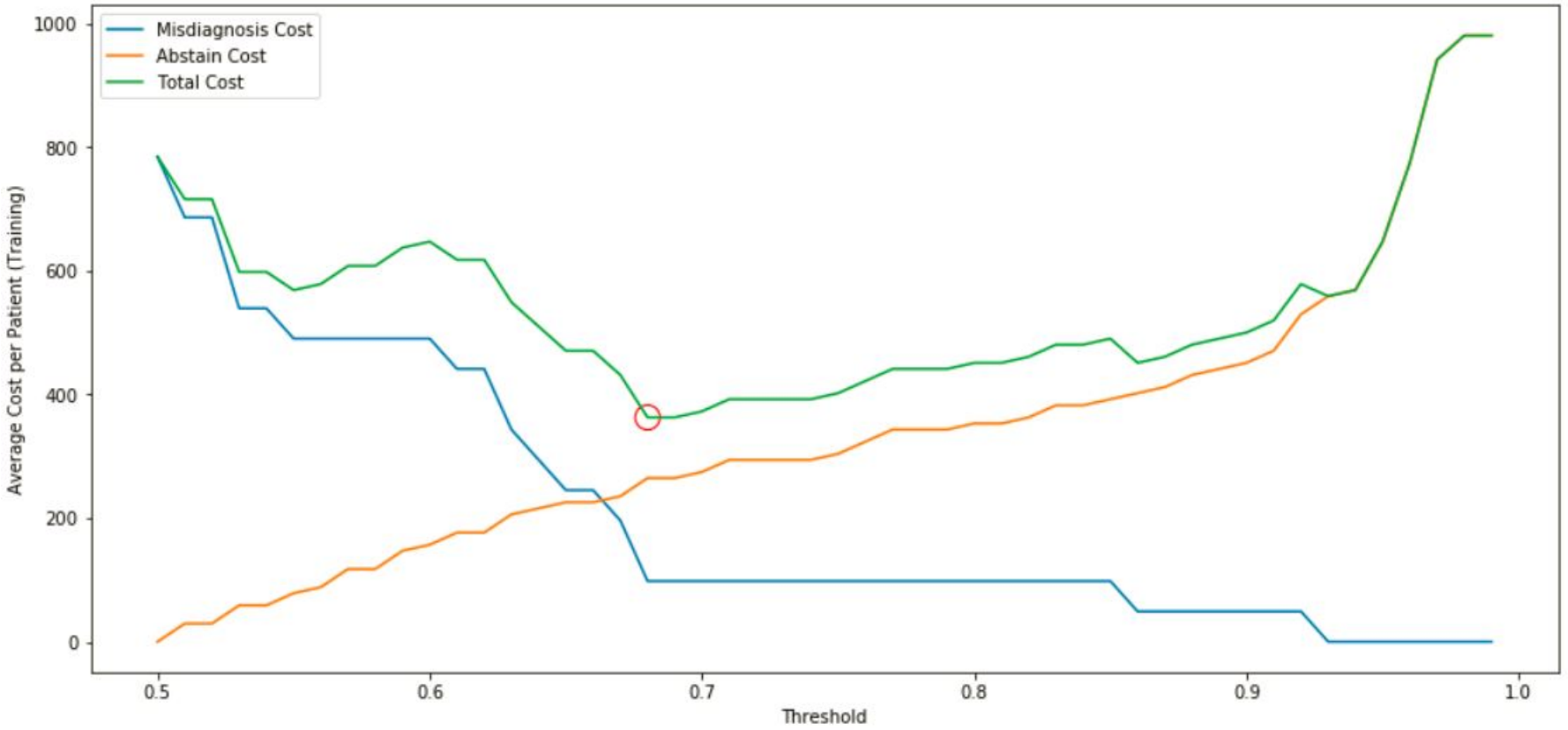
- With OvR default threshold th = 0.5 (A = 0)
  - $T = \frac{5000M+1000A}{N} = \frac{5000M}{N} = \frac{5000*16}{102} = \$784.3$
- With threshold th = 1 (A=N)
  - T = $1000

y_train_ovr	1.0	2.0	3.0	All
y_train				
1.0	69	4	0	73
2.0	9	9	0	18
3.0	3	0	8	11
All	81	13	8	102

Idea:

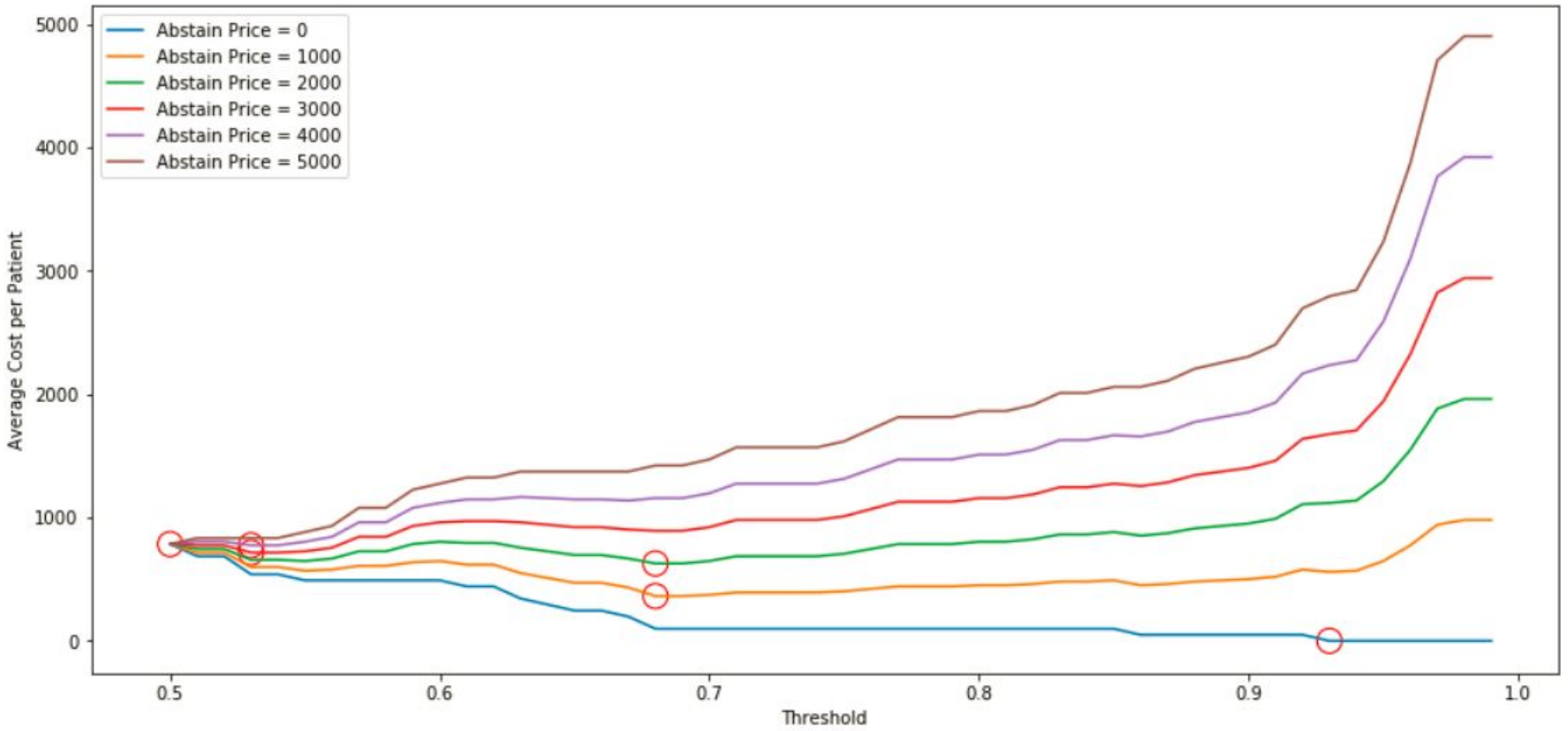
- Set new threshold th (if p > th predict; if p < th abstain)
  - th ↑ → M ↓ and A ↑
  - th ↓ → M ↑ and A ↓
- Optimal threshold th ≠ 0.5 to minimize cost T
- Sweep threshold th to get minimum cost T

THRESHOLD ANALYSIS



- Minimum total cost T = $362.75 (Training Set)
- New Threshold th = 0.68 ≠ 0.5 (default)
- Minimum total cost T = $575.22 (Testing Set)

METHOD ROBUSTNESS



- Always found the optimal threshold
- Verification:
  - When abstain price is low, abstain to predict cost leas – high threshold
  - When abstain price is high, abstain to predict cost more – low threshold