Trading off Precision and Recall

Handling Skewed Data

Advice for Applying Machine Learning:

Trading off precision and recall

Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$

Predict 0 if $h_\theta(x) < 0.5$

precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

Trading off precision and recall

- Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$

Predict 0 if $h_\theta(x) < 0.5$

Suppose we want to predict $y = 1$ (cancer) only if very confident.

- precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$
- recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

Trading off precision and recall

- Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ 0.7

Predict 0 if $h_\theta(x) < 0.5$ 0.7

Suppose we want to predict $y = 1$ (cancer) only if very confident.

- precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

- recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

Trading off precision and recall

→ precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

→ recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

→ Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ 0.7

Predict 0 if $h_\theta(x) < 0.5$ 0.7

Suppose we want to predict $y = 1$ (cancer) only if very confident.

→ Higher precision,

Trading off precision and recall

- Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ 0.7

Predict 0 if $h_\theta(x) < 0.5$ 0.7

Suppose we want to predict $y = 1$ (cancer) only if very confident.

→ Higher precision, lower recall.

→ precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

→ recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

Trading off precision and recall

→ precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

→ recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

→ Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ 0.7 0.9

Predict 0 if $h_\theta(x) < 0.5$ 0.7 0.9

Suppose we want to predict $y = 1$ (cancer) only if very confident.

→ Higher precision, lower recall.

Trading off precision and recall

→ precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

→ recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

→ Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ 0.7 0.9

Predict 0 if $h_\theta(x) < 0.5$ 0.7 0.9

Suppose we want to predict $y = 1$ (cancer) only if very confident.

→ Higher precision, lower recall.

Suppose we want to avoid missing too many cases of cancer (avoid false negatives).

Trading off precision and recall

→ precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

→ recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

→ Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ 0.7 0.9 0.3

Predict 0 if $h_\theta(x) < 0.5$ 0.7 0.9 0.3

Suppose we want to predict $y = 1$ (cancer) only if very confident.

→ Higher precision, lower recall.

Suppose we want to avoid missing too many cases of cancer (avoid false negatives).

Trading off precision and recall

→ precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

→ recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

→ Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ ~~0.5~~ ~~0.7~~ ~~0.9~~ 0.3

Predict 0 if $h_\theta(x) < 0.5$ ~~0.5~~ ~~0.7~~ ~~0.9~~ 0.3

Suppose we want to predict $y = 1$ (cancer) only if very confident.

→ Higher precision, lower recall.

Suppose we want to avoid missing too many cases of cancer (avoid false negatives).

→ Higher recall, lower precision.

Trading off precision and recall

→ precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

→ recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

→ Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ ~~0.7~~ ~~0.9~~ 0.3

Predict 0 if $h_\theta(x) < 0.5$ ~~0.7~~ ~~0.9~~ 0.3

→ Suppose we want to predict $y = 1$ (cancer) only if very confident.

→ Higher precision, lower recall.

→ Suppose we want to avoid missing too many cases of cancer (avoid false negatives).

→ Higher recall, lower precision.

More generally: Predict 1 if $h_\theta(x) \geq$ threshold.

Trading off precision and recall

→ precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

→ recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

→ Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ 0.7 0.9 0.3

Predict 0 if $h_\theta(x) < 0.5$ 0.7 0.9 0.3

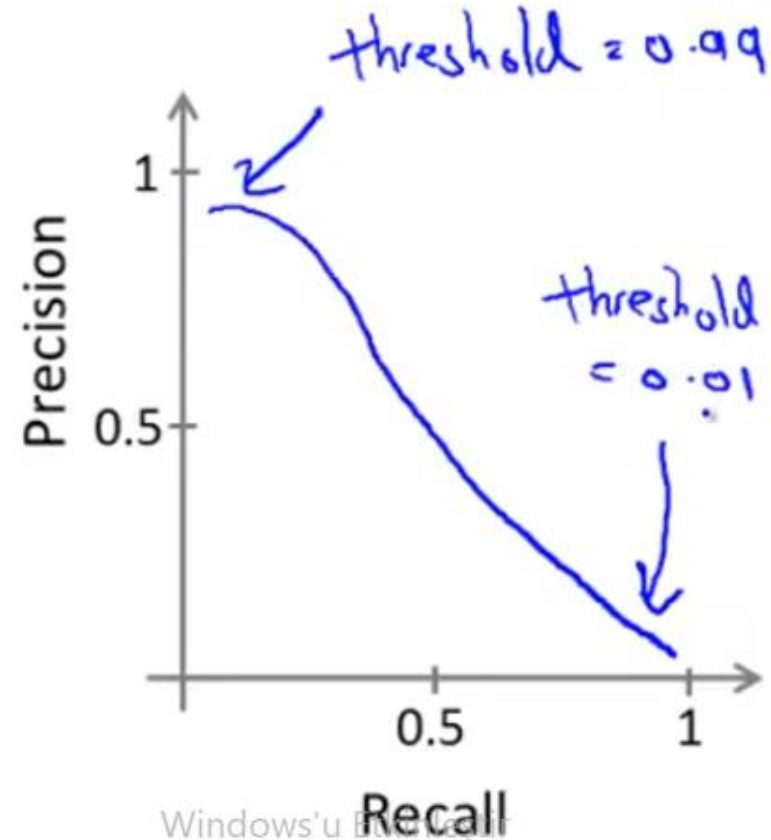
→ Suppose we want to predict $y = 1$ (cancer) only if very confident.

→ Higher precision, lower recall.

→ Suppose we want to avoid missing too many cases of cancer (avoid false negatives).

→ Higher recall, lower precision.

More generally: Predict 1 if $h_\theta(x) \geq$ threshold.



Trading off precision and recall

→ precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$

→ recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$

→ Logistic regression: $0 \leq h_\theta(x) \leq 1$

Predict 1 if $h_\theta(x) \geq 0.5$ 0.7 0.9 0.3

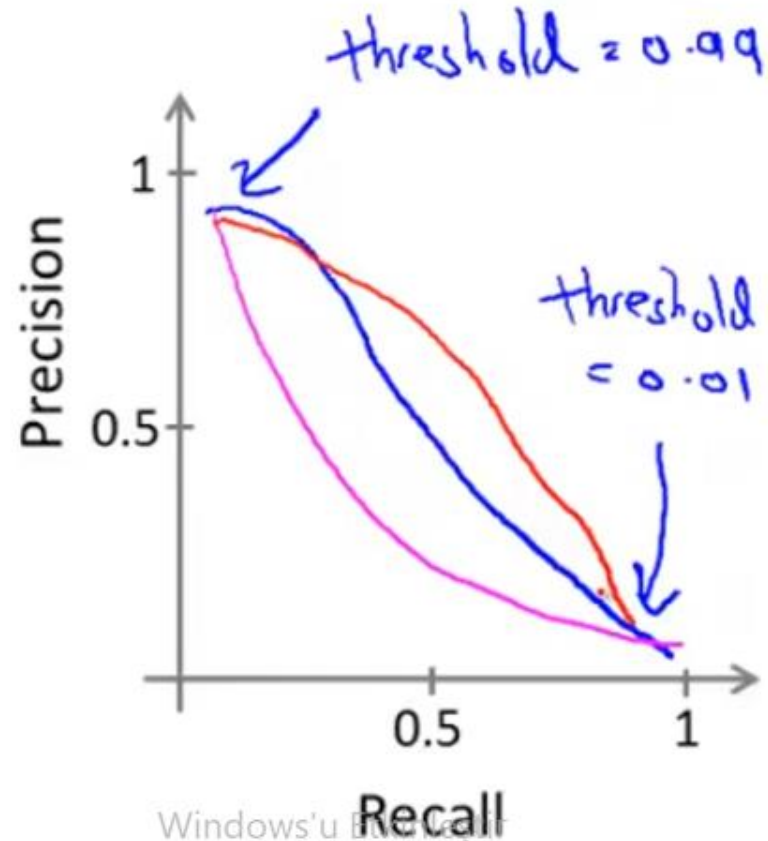
Predict 0 if $h_\theta(x) < 0.5$ 0.7 0.9 0.3

→ Suppose we want to predict $y = 1$ (cancer) only if very confident.

→ Higher precision, lower recall.

→ Suppose we want to avoid missing too many cases of cancer (avoid false negatives).

→ Higher recall, lower precision.



More generally: Predict 1 if $h_\theta(x) \geq$ threshold.

F_1 Score (F score)

How to compare precision/recall numbers?

	Precision(P)	Recall (R)
Algorithm 1	0.5	0.4
Algorithm 2	0.7	0.1
Algorithm 3	0.02	1.0

F_1 Score (F score)

How to compare precision/recall numbers?

	Precision(P)	Recall (R)	Average
Algorithm 1	0.5	0.4	0.45
Algorithm 2	0.7	0.1	0.4
Algorithm 3	0.02	1.0	0.51

Average: $\frac{P+R}{2}$

F_1 Score (F score)

How to compare precision/recall numbers?

	Precision(P)	Recall (R)	Average
Algorithm 1	0.5	0.4	0.45
Algorithm 2	0.7	0.1	0.4
Algorithm 3	0.02	1.0	0.51

Predict y=1 all the time

Average: $\frac{P+R}{2}$

F_1 Score (F score)

How to compare precision/recall numbers?

	Precision(P)	Recall (R)	Average
Algorithm 1	0.5	0.4	0.45
Algorithm 2	0.7	0.1	0.4
Algorithm 3	0.02	1.0	0.51

Average: $\frac{P+R}{2}$

Predict y=1 all the time

F_1 Score (F score)

How to compare precision/recall numbers?

	Precision(P)	Recall (R)	Average	F_1 Score
Algorithm 1	0.5	0.4	0.45	0.444
Algorithm 2	0.7	0.1	0.4	0.175
Algorithm 3	0.02	1.0	0.51	0.0392

Predict y=1 all the time

Average: $\frac{P+R}{2}$

F_1 Score: $2\frac{PR}{P+R}$