Metrics

AUROC - area under receiver operating characteristic

Asks: what is the probability that a random true positive will be ranked higher than a random true negative? *Measures ranking at all thresholds*

	Predict hit	Predict no hit
ASMS hit	TP	FN
ASMS no hit	FP	TN

$$\frac{TP}{TP + FN}$$

True positive rate aka recall

$$\frac{FP}{FP + TN}$$

False positive rate

AUROC - area under receiver operating characteristic

Asks: what is the probability that a random true positive will be ranked higher than a random true negative? *Measures ranking at all thresholds*

	Predict hit	Predict no hit
ASMS hit	TP	FN
ASMS no hit	FP	TN

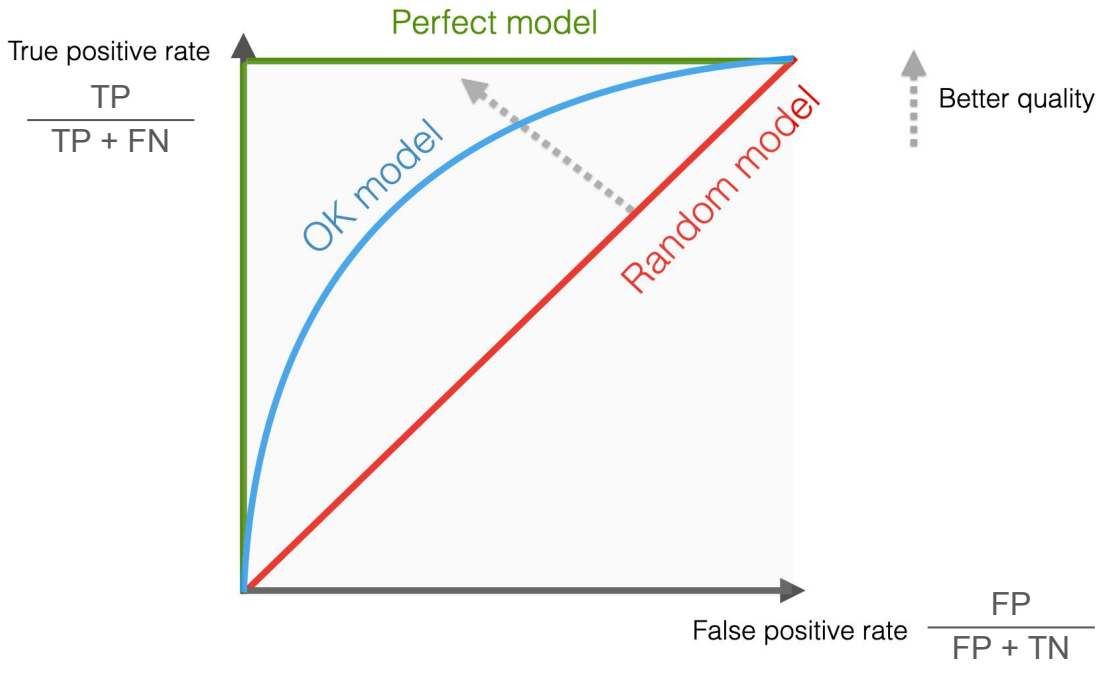
Perfect model	Predict hit	Predict no hit
ASMS hit	**100%**	**0%**
ASMS no hit	**0%**	**100%**

Random model	Predict hit	Predict no hit
ASMS hit	**50%**	**50%**
ASMS no hit	**50%**	**50%**

Row percentages

AUROC - area under receiver operating characteristic

	Predict hit	Predict no hit
ASMS hit	TP	FN
ASMS no hit	FP	TN



https://www.evidentlyai.com/classification-metrics/explain-roc-curve

AUPRC - area under precision recall curve

Asks: how hit-rich are my top ranked predictions? *Measures expected precision at all thresholds*

	Predict hit	Predict no hit
ASMS hit	TP	FN
ASMS no hit	FP	TN

$$\frac{TP}{TP + FN}$$

True positive rate aka recall

$$\frac{TP}{TP + FP}$$

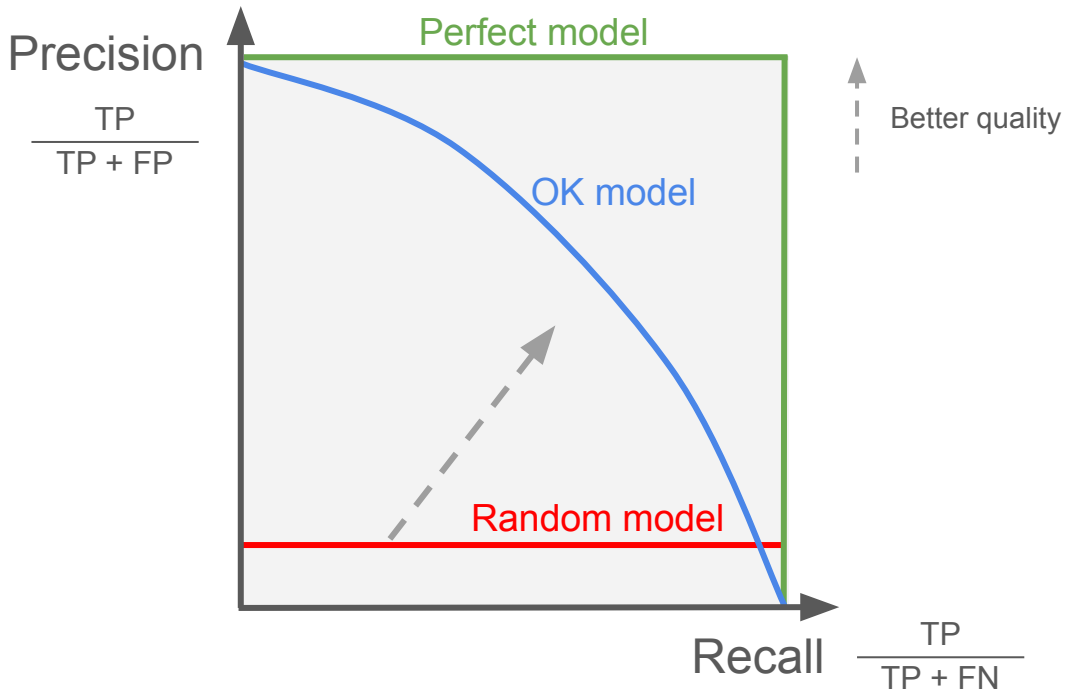
Precision

Interested in a row % and a column %

AUPRC - area under precision recall curve

Asks: how hit-rich are my top ranked predictions? *Measures expected precision at all thresholds*

	Predict hit	Predict no hit
ASMS hit	TP	FN
ASMS no hit	FP	TN



We care most about the top ranked molecules *not necessarily performance at all thresholds*

Predictions from model

Molecule	Predicted probability
E	0.65
B	0.40
F	0.20
A	0.12
C	0.03
D	0.01

Test labels from ASMS

Molecule	ASMS Hit (ground truth)
A	1
B	1
C	0
D	1
E	1
F	0

Hits at 3 *How many TP are in top 3?*

Molecule	Predicted probability
E	0.65
B	0.40
F	0.20
A	0.12
C	0.03
D	0.01

Molecule	ASMS Hit (ground truth)
A	1
B	1
C	0
D	1
E	1
F	0

Hits at 3 = 2

Molecule	Predicted probability	
E	0.65	✅
B	0.40	✅
F	0.20	
A	0.12	
C	0.03	
D	0.01	

Molecule	ASMS Hit (ground truth)
A	1
B	1
C	0
D	1
E	1
F	0

Precision at 3 *what % of top 3 are TP?*

Molecule	Predicted probability
E	0.65
B	0.40
F	0.20
A	0.12
C	0.03
D	0.01

Molecule	ASMS Hit (ground truth)
A	1
B	1
C	0
D	1
E	1
F	0

Precision at 3 = 2/3 = 0.66

Molecule	Predicted probability	
E	0.65	✅
B	0.40	✅
F	0.20	❌
A	0.12	
C	0.03	
D	0.01	

Molecule	ASMS Hit (ground truth)
A	1
B	1
C	0
D	1
E	1
F	0

Recall at 3 *what % of TP are in the top 3?*

Molecule	Predicted probability
E	0.65
B	0.40
F	0.20
A	0.12
C	0.03
D	0.01

Molecule	ASMS Hit (ground truth)
A	1
B	1
C	0
D	1
E	1
F	0

Recall at 3 = 2/4 = 0.5

Molecule	Predicted probability	
E	0.65	✅
B	0.40	✅
F	0.20	
A	0.12	❌
C	0.03	
D	0.01	❌

Molecule	ASMS Hit (ground truth)
A	1
B	1
C	0
D	1
E	1
F	0

Why not threshold? It's too stringent

Want to get credit for ranking B highly!

Molecule	Predicted probability	Pred proba >0.5	
E	0.65	1	✅
B	0.40	0	❌
F	0.20	0	
A	0.12	0	❌
C	0.03	0	
D	0.01	0	❌

Molecule	ASMS Hit (ground truth)
A	1
B	1
C	0
D	1
E	1
F	0

Summary

- **AUROC** measures ranking ability at all thresholds
- **AUPRC** measures expected precision at all thresholds
- **Hits @ K** measures number of True Positives in top K
- **Precision @ K** measures percentage of top K which are True Positives
- **Recall @ K** measures percentage of True Positives which are in the top K