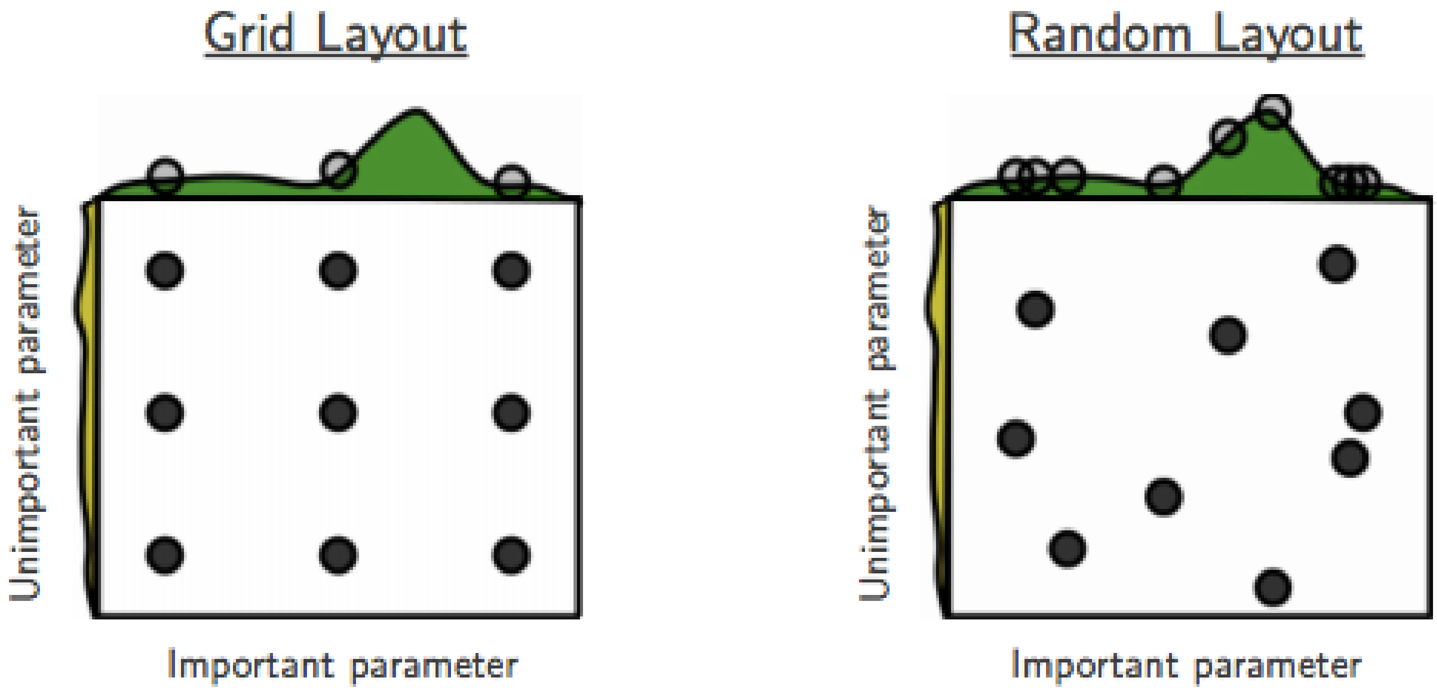
Adversarial examples, face verification

Bálint Ármin Pataki

Hyperparameter search

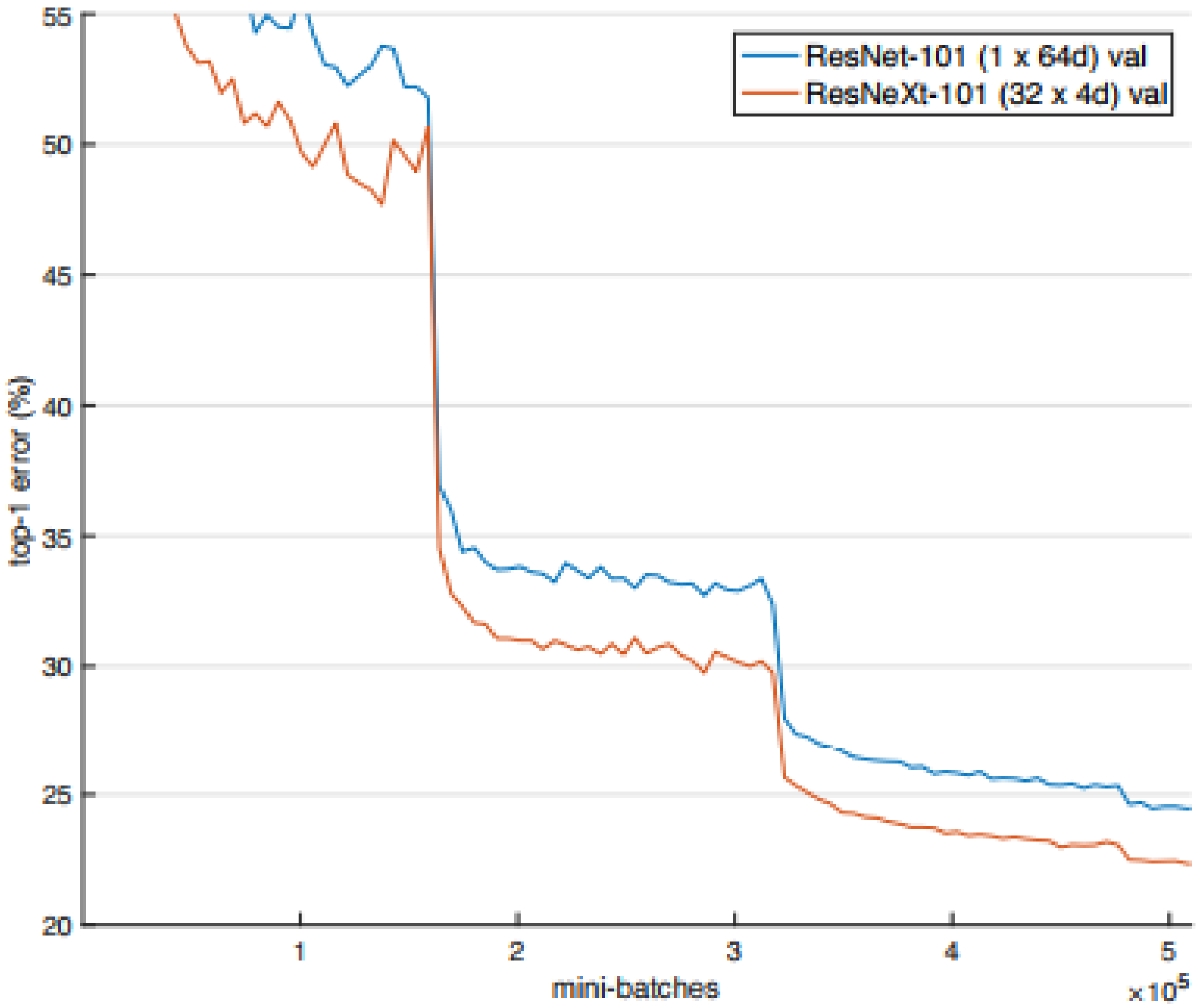
Hyperparameter: the one are tuned by hand
Parameter: the ones are tuned by the optimization algorithm

Diagrammatic representation of Grid Search by Bergstra & Bengio



Learning rate is special

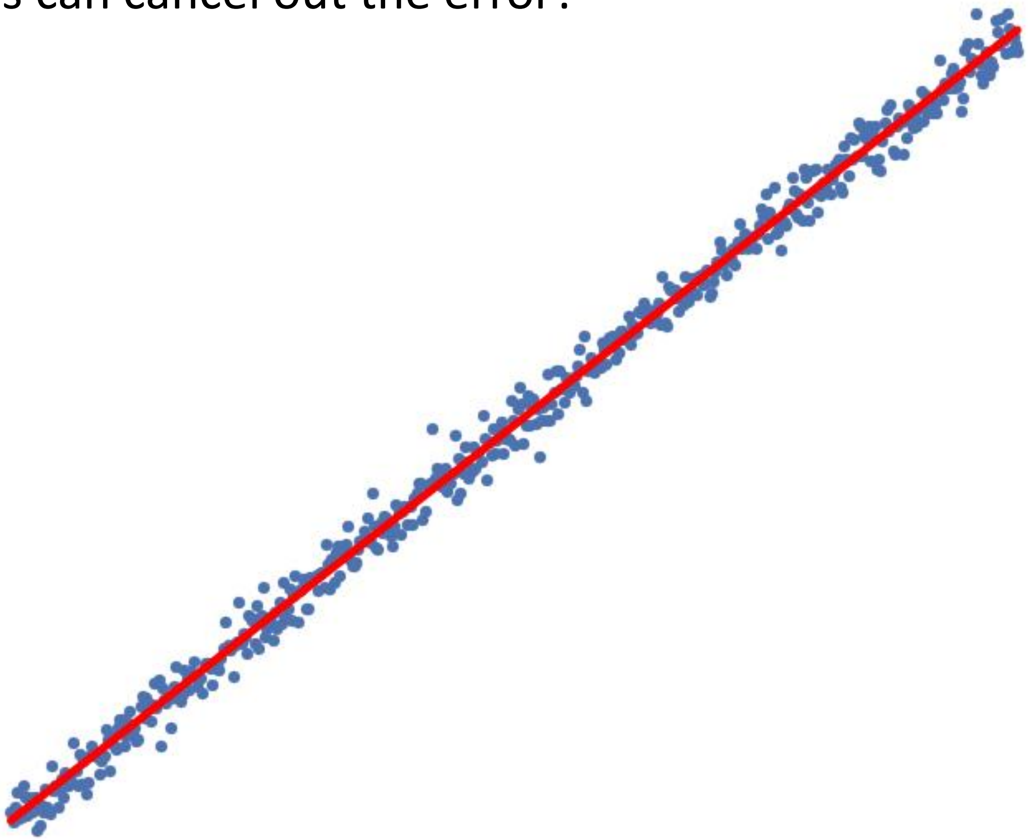
Pre-defined schedule vs baby-sitting



[Xie: Aggregated Residual Transformations for Deep Neural Networks arXiv:1611.05431v2]

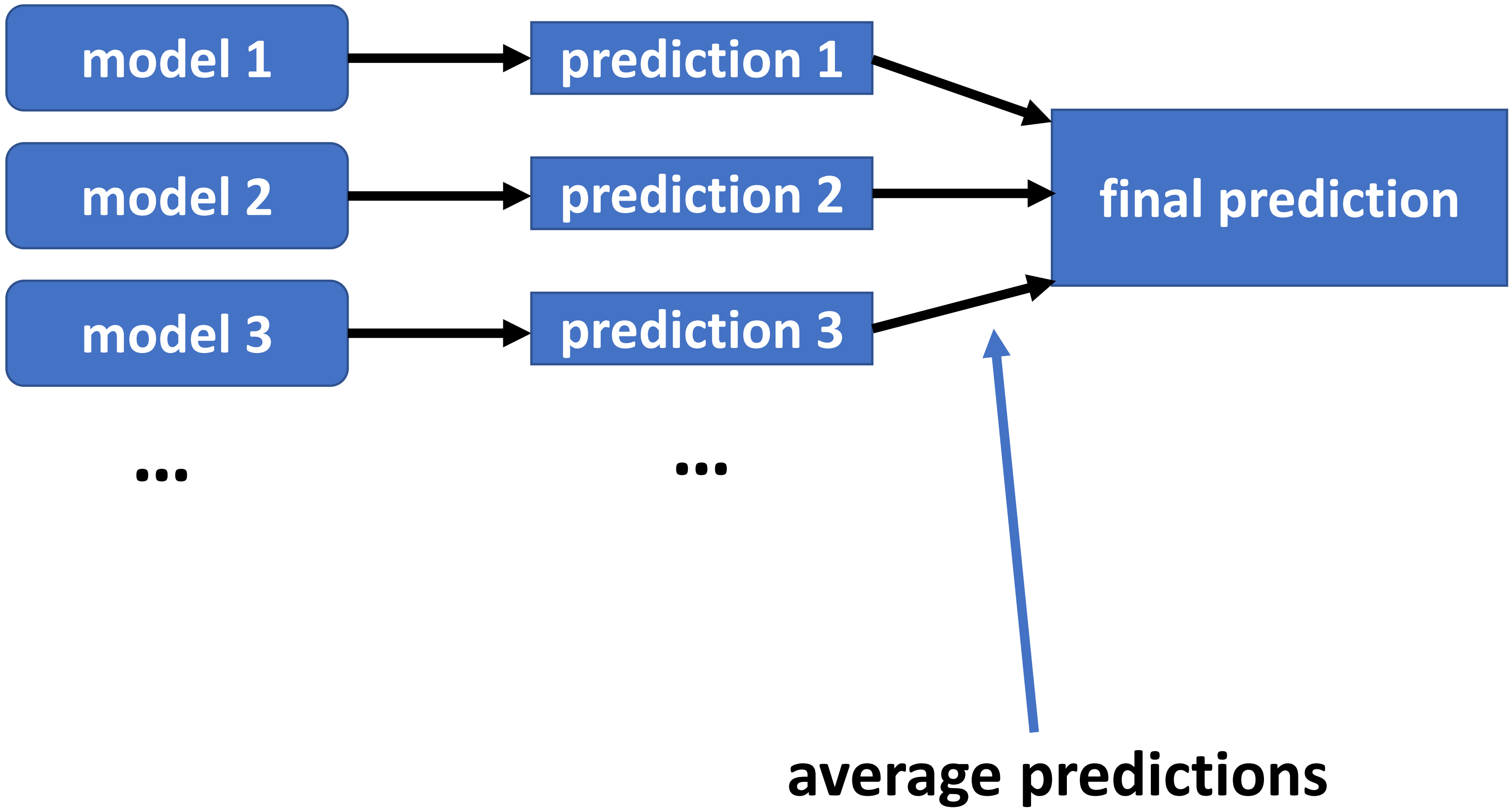
Ensemble - motivation

What if the error is random, but model dependent?
→training different models can cancel out the error?



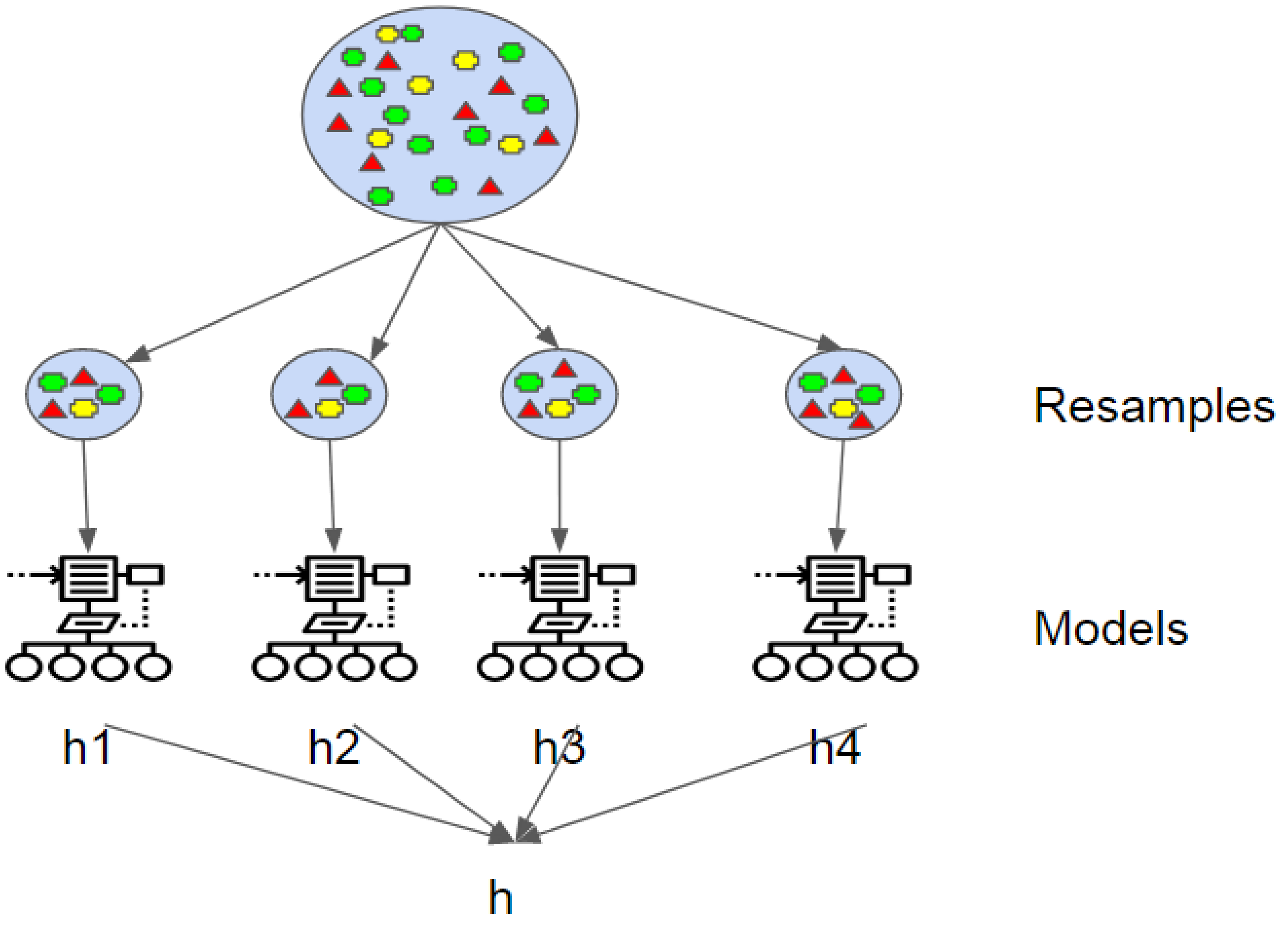
At Kaggle always an ensemble wins.
Often it has no practical relevance, but it can increase the score with epsilon.

Ensemble



Special case: Bagging (Bootstap AGGregatING)
- same models trained on subset of train data

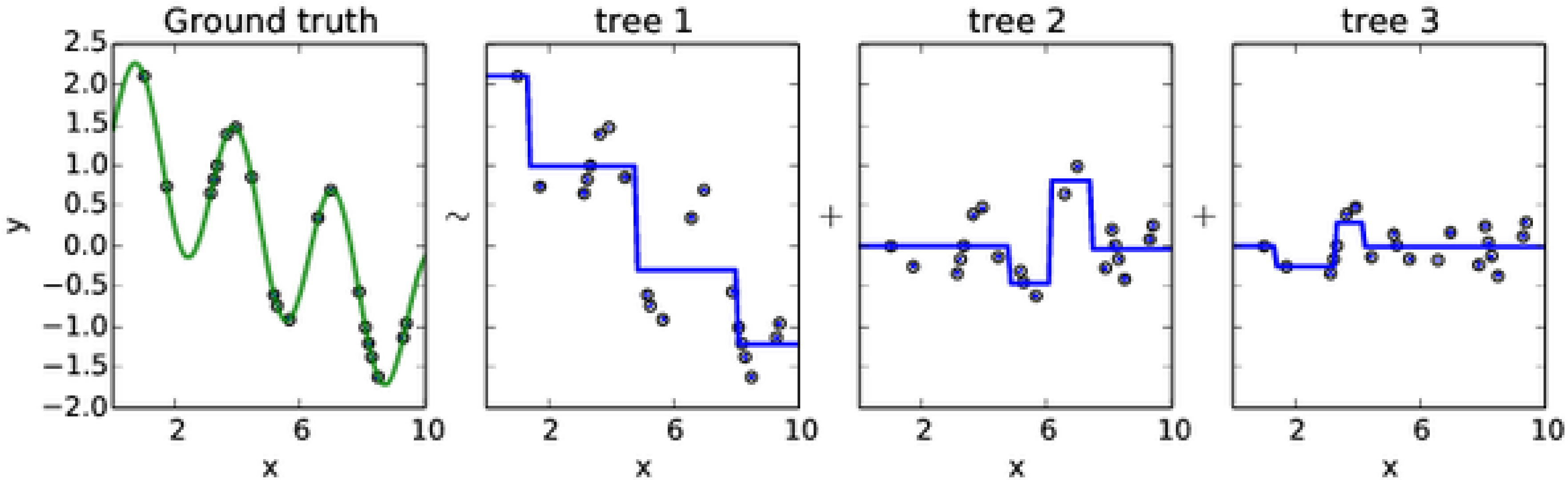
Bagging



https://medium.com/@SeattleDataGuy/how-to-develop-a-robust-algorithm-c38e08f32201

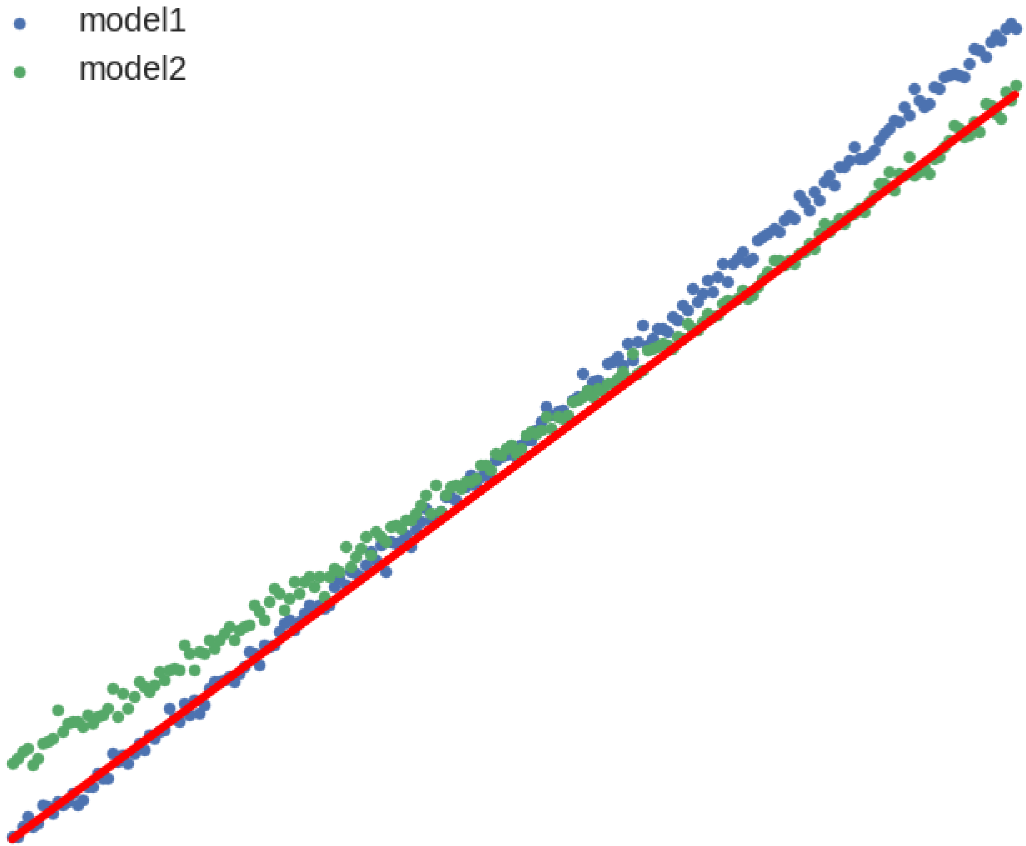
Gradient boosting

Model 1 fits the original data.
Model 2 fits original data – model 1 prediction = the residual
Model 3 fits original data – model 1 prediction – model 2 prediction
…



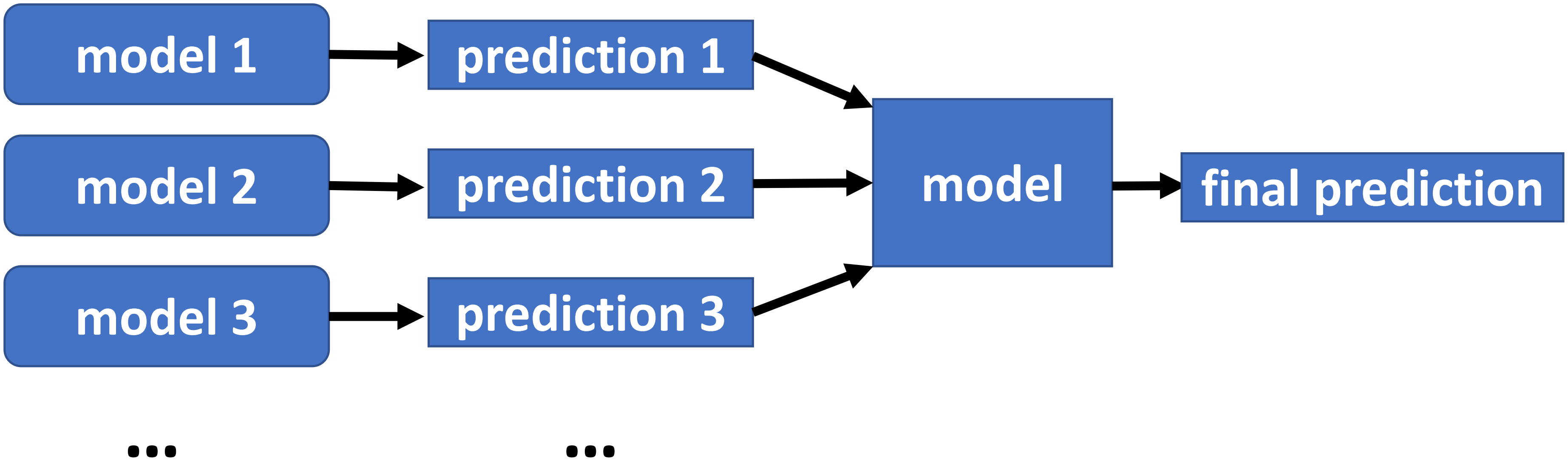
https://www.quora.com/How-would-you-explain-gradient-boosting-machine-learning-technique-in-no-more-than-300-words-to-non-science-major-college-students

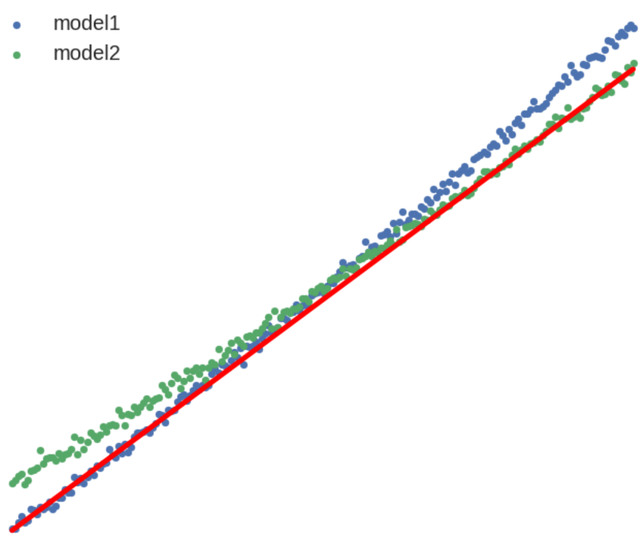
Stacking - motivation



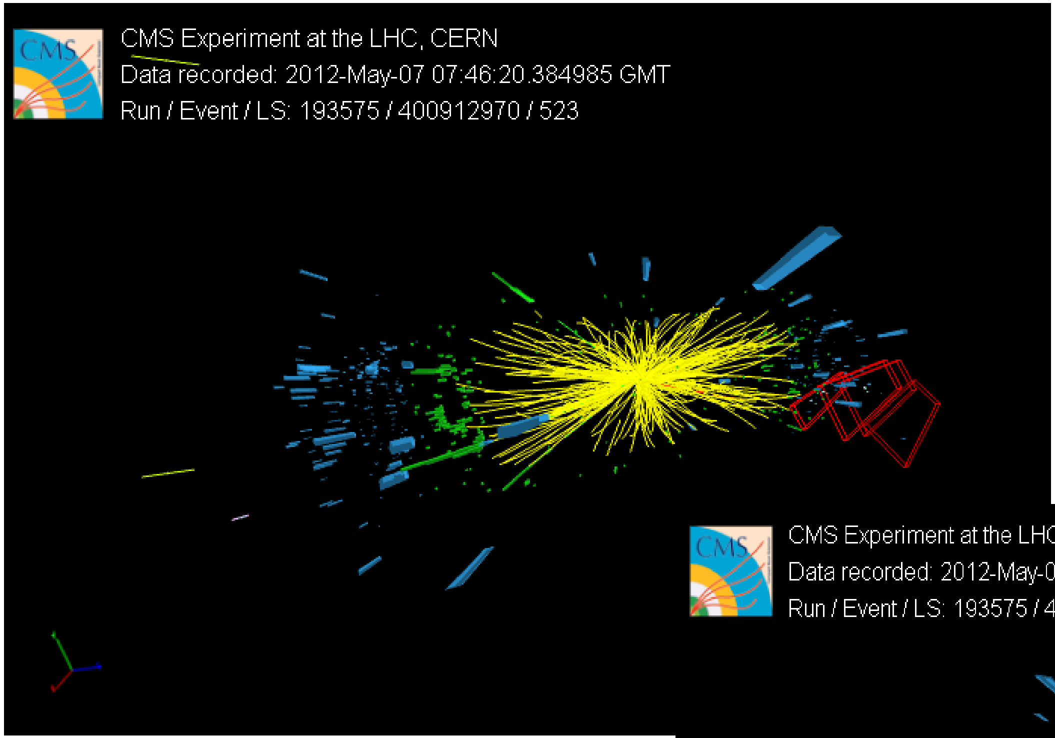
Model 1 is string for smaller X, model 2 is for larger X.
Would be great to combine them!

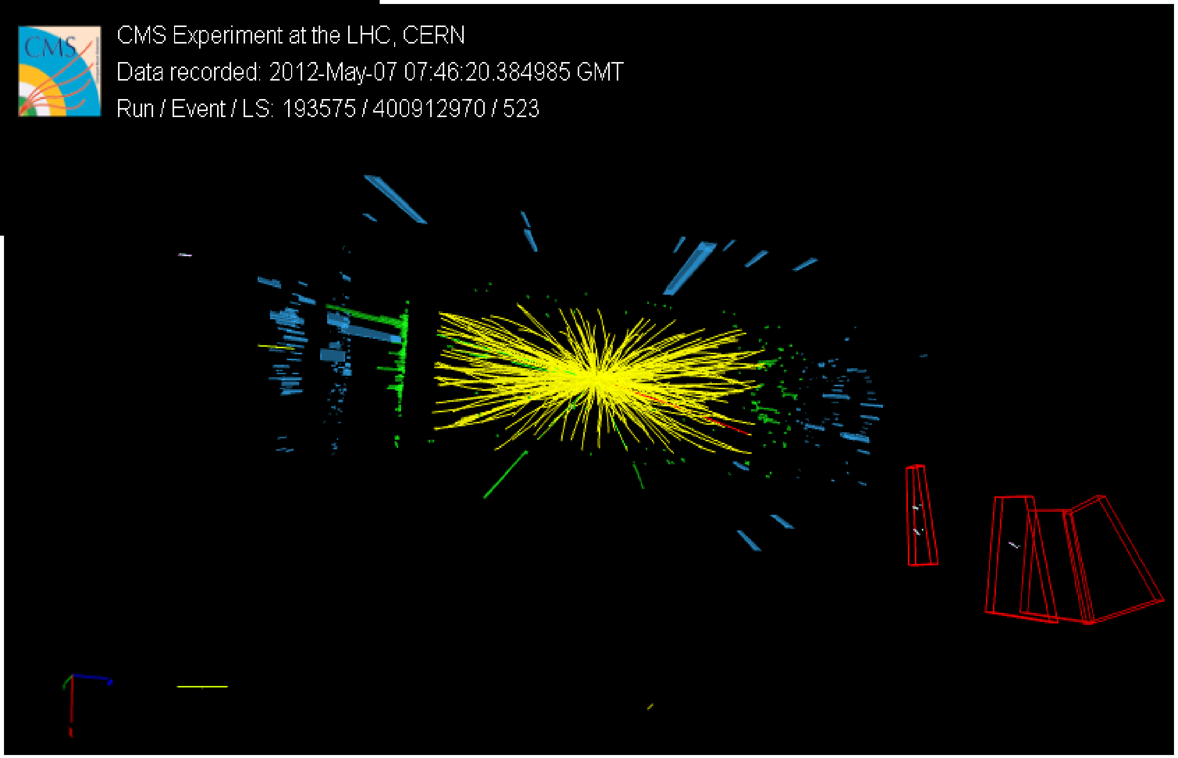
Stacking


model 1
model 2
model 3
...
prediction 1
prediction 2
prediction 3
...
model
final prediction


model1
model2

Proper validation strategy





- **diff. view of same event**
- **diff. measurement of same person**
- **time series**
- **...**

DEMO notebook

Other PPT

http://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/presentation.pptx

Importance of the correct metric

https://www.walesonline.co.uk/news/wales-news/facial-recognition-wrongly-identified-2000-14619145

Facial recognition software wrongly identified more than 2,000 people as potential criminals as police patrolled the Champions League final in Cardiff .

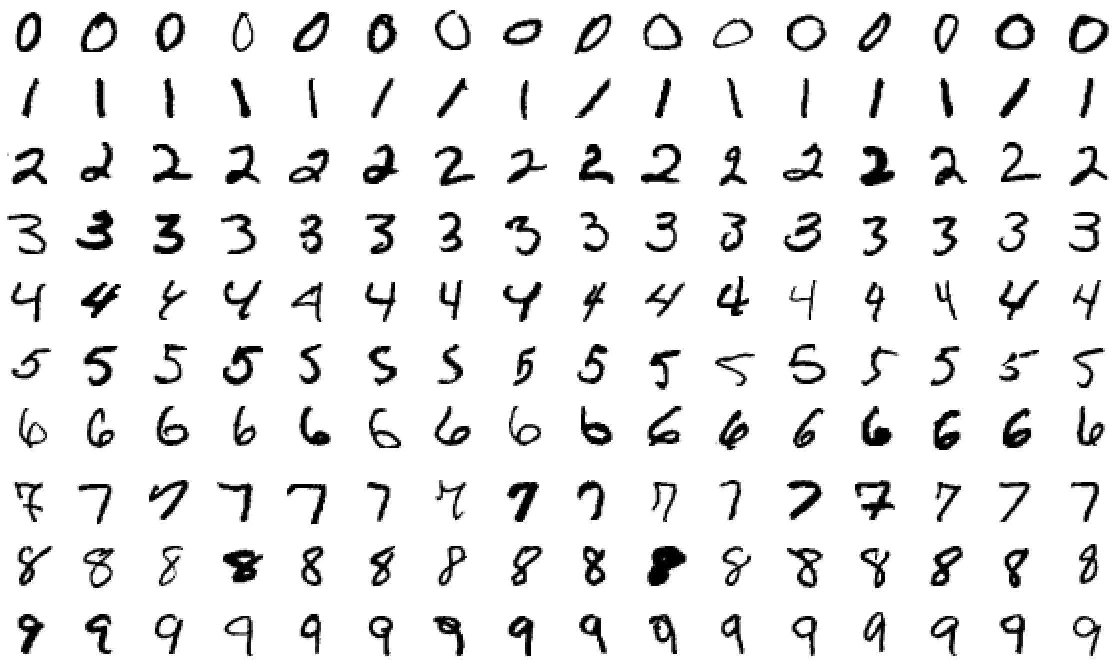
The technology provided hundreds of "false positives" wrongly marking out innocent people as possible troublemakers when an estimated 170,000 people descended on the city for the showpiece match between Real Madrid and Juventus.

A South Wales Police spokesman admitted "no facial recognition system is 100% accurate under all conditions" but added that in the months since it was first deployed "no-one has been arrested where a 'false positive alert' has occurred and no members of the public have complained".

Data published by the force showed police covering the Champions League final at the Principality Stadium on June 3 last year were alerted to 2,470 potential matches with custody pictures by the facial recognition programme.

But of these 92% – a total if 2,297 – were incorrect, with just 173 providing 'true positive alerts'.

Importance of the correct metric



MNIST

- one vs all classifier (zero or not zero)
- metric: accuracy
- 90% accuracy. Is it good?

Instead of accuracy

Positive = predicted class

True = prediction is correct

True Positive (TP)	**True Negative (TN)**
False Positive (FP)	**False Negative (FN)**

Is this digit 0 (actually it is)?
Prediction: yes
→True positive

Is this digit 0 (actually it is not)?
Prediction: no
→True negative

Is this digit 0 (actually it is not)?
Prediction: yes
→False positive

Is this digit 0 (actually it is)?
Prediction: no
→False negative

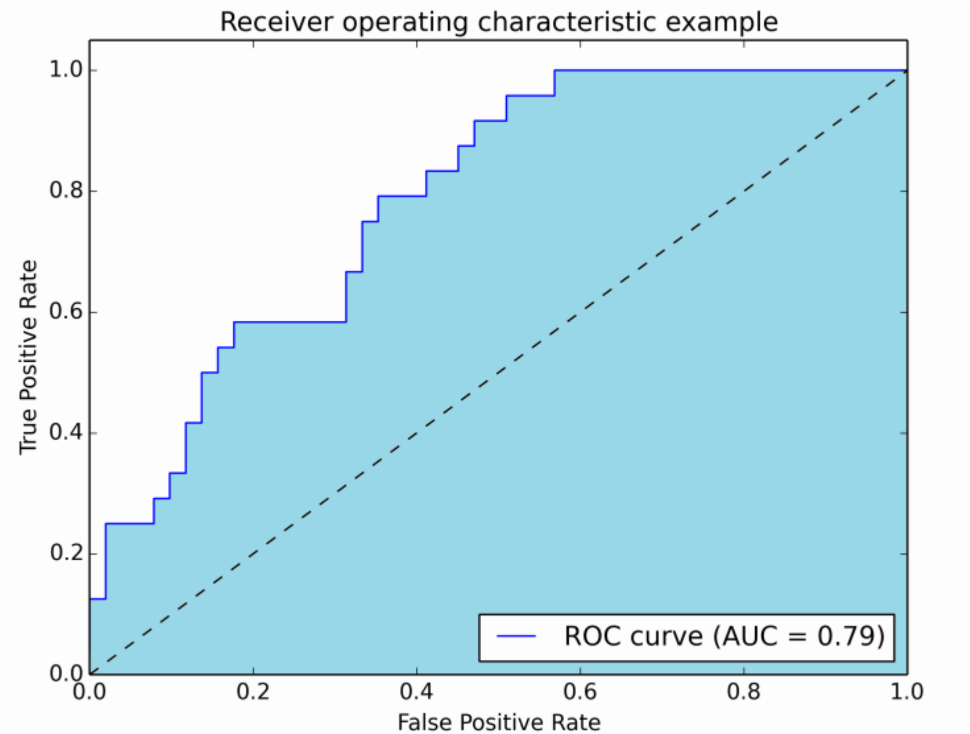
Instead of accuracy – AUC (ROC)

You predict a probability of being positive.
Then a threshold is applied (eq 50%) and is the probability is above, then prediction is positive

For different thesholds there is different prediction.

For different tasks you want different goals:
- identification of criminals: avoid False Negative
- unlock your phone with face recog: avoid False Positive

http://www.navan.name/roc/



https://stats.stackexchange.com/questions/132777/what-does-auc-stand-for-and-what-is-it

Instead of accuracy – many more

Problem dependent. Objectives:
- easy to understand/interpret
- significantly better model should have significantly better score
- cover your exact need (as possible)

		True condition				
	Total population	Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**, Power	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$	F_1 score = $\frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	True negative rate (TNR), Specificity (SPC) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$		

https://en.wikipedia.org/wiki/Precision_and_recall

Instead of accuracy – many more and a few more

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{P} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

specificity or true negative rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{N} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$$

precision or positive predictive value (PPV)

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

negative predictive value (NPV)

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}}$$

miss rate or false negative rate (FNR)

$$\mathrm{FNR} = \frac{\mathrm{FN}}{P} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TP}} = 1-\mathrm{TPR}$$

fall-out or false positive rate (FPR)

$$\mathrm{FPR} = \frac{\mathrm{FP}}{N} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = 1-\mathrm{TNR}$$

false discovery rate (FDR)

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TP}} = 1-\mathrm{PPV}$$

false omission rate (FOR)

$$\mathrm{FOR} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TN}} = 1-\mathrm{NPV}$$

accuracy (ACC)

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{P+N} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

F1 score

is the harmonic mean of precision and sensitivity

$$F_1 = 2 \cdot \frac{\mathrm{PPV}\cdot\mathrm{TPR}}{\mathrm{PPV}+\mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

Matthews correlation coefficient (MCC)

$$\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

Informedness or Bookmaker Informedness (BM)

$$\mathrm{BM} = \mathrm{TPR}+\mathrm{TNR}-1$$

Markedness (MK)

$$\mathrm{MK} = \mathrm{PPV}+\mathrm{NPV}-1$$

Sources: Fawcett (2006), Powers (2011), and Ting (2011) [4] [1] [5]

https://en.wikipedia.org/wiki/Precision_and_recall

Importance of the correct metric

https://www.walesonline.co.uk/news/wales-news/facial-recognition-wrongly-identified-2000-14619145

Facial recognition software wrongly identified more than 2,000 people as potential criminals as police patrolled the **Champions League final** in **Cardiff** .

The technology provided hundreds of "false positives" wrongly marking out innocent people as possible troublemakers when an estimated 170,000 people descended on the city for the showpiece match between Real Madrid and Juventus.

A **South Wales Police** spokesman admitted "no facial recognition system is 100% accurate under all conditions" but added that in the months since it was first deployed "no-one has been arrested where a 'false positive alert' has occurred and no members of the public have complained".

Data **published by the force** showed police covering the Champions League final at the Principality Stadium on June 3 last year were alerted to 2,470 potential matches with custody pictures by the facial recognition programme.

But of these 92% – a total if 2,297 – were incorrect, with just 173 providing 'true positive alerts'.

Links

https://blog.openai.com/adversarial-example-research/

https://blog.openai.com/robust-adversarial-inputs/

http://www.navan.name/roc/

http://arogozhnikov.github.io/2016/07/05/gradient_boosting_playground.html

http://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/presentation.pptx