CALVIN INSTITUTE OF TECHNOLOGY

God's People for God's Glory

Classification

Hendrik Santoso Sugiarto

IBDA2032 – *Artificial Intelligence*

Capaian Pembelajaran

- Konsep Klasifikasi
- Klasifikasi Biner
- Klasifikasi Multi-Kelas
- Klasifikasi Multi-Label

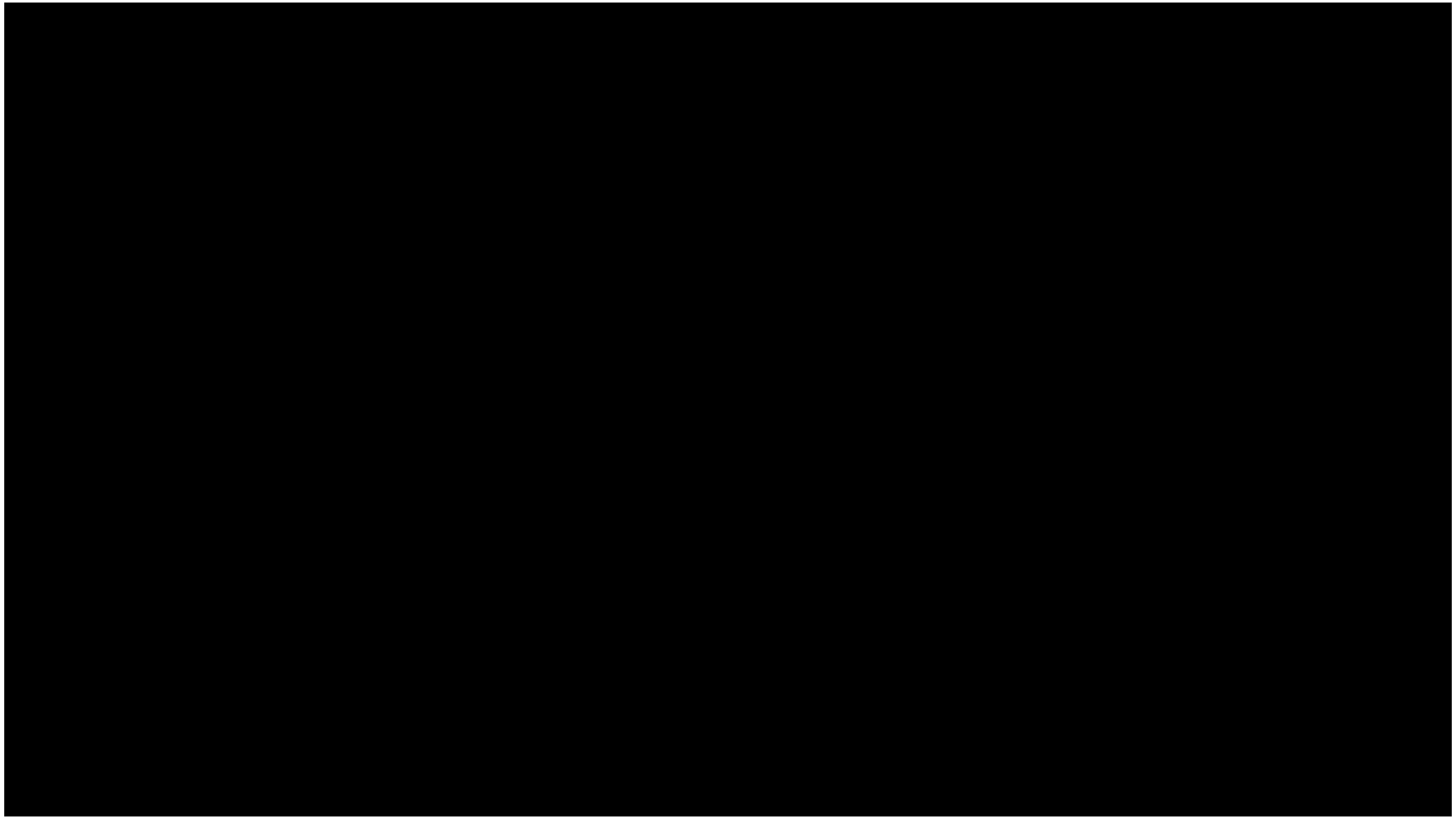
Klasifikasi

Supervised Learning

- Bentuk formal:
  - Input: $\boldsymbol{x} \in \mathcal{X} \in \mathbb{R}^n$
  - Output: $y \in \mathcal{Y} \in \begin{cases} \mathbb{R} \to regresi \quad \text{Real} \\ \{+1, -1\} \to klasifikasi\ biner \quad / \ (1, 0) \\ \{1,2, \ldots, K\} \to klasifikasi\ multikelas \end{cases}$
  - Target function: $f: \mathcal{X} \to \mathcal{Y}\ (unknown)$
  - Training data: $D = \{(\boldsymbol{x}^{(1)}, y^{(1)}), (\boldsymbol{x}^{(2)}, y^{(2)}), \ldots, (\boldsymbol{x}^{(m)}, y^{(m)})\}$
  - Hypothesis: $h: \mathcal{X} \to \mathcal{Y}$
  - Hypothesis space: $h \in \mathcal{H}$

CALVIN INSTITUTE OF TECHNOLOGY

Klasifikasi Satu Kelas



https://www.youtube.com/watch?v=bDg76h9-jnQ Diakses pada tanggal 7 Maret 2021

Klasifikasi Multi Kelas

https://www.youtube.com/watch?v=_pm7IrGmalQ Diakses pada tanggal 7 Maret 2021

Klasifikasi

- Proses mengelompokan sekumpulan data ke dalam kategori yang telah ditentukan sebelumnya
- Tipe klasifikasi:
  - Klasifikasi biner (ya/tidak, laki-laki/perempuan)
  - Klasifikasi multi-kelas (identifikasi suatu gambar angka)
  - Klasifikasi multi-label (identifikasi beberapa objek di satu gambar)
  - Klasifikasi multi-output (identifikasi objek dan lokasinya)

Uji Pemahaman

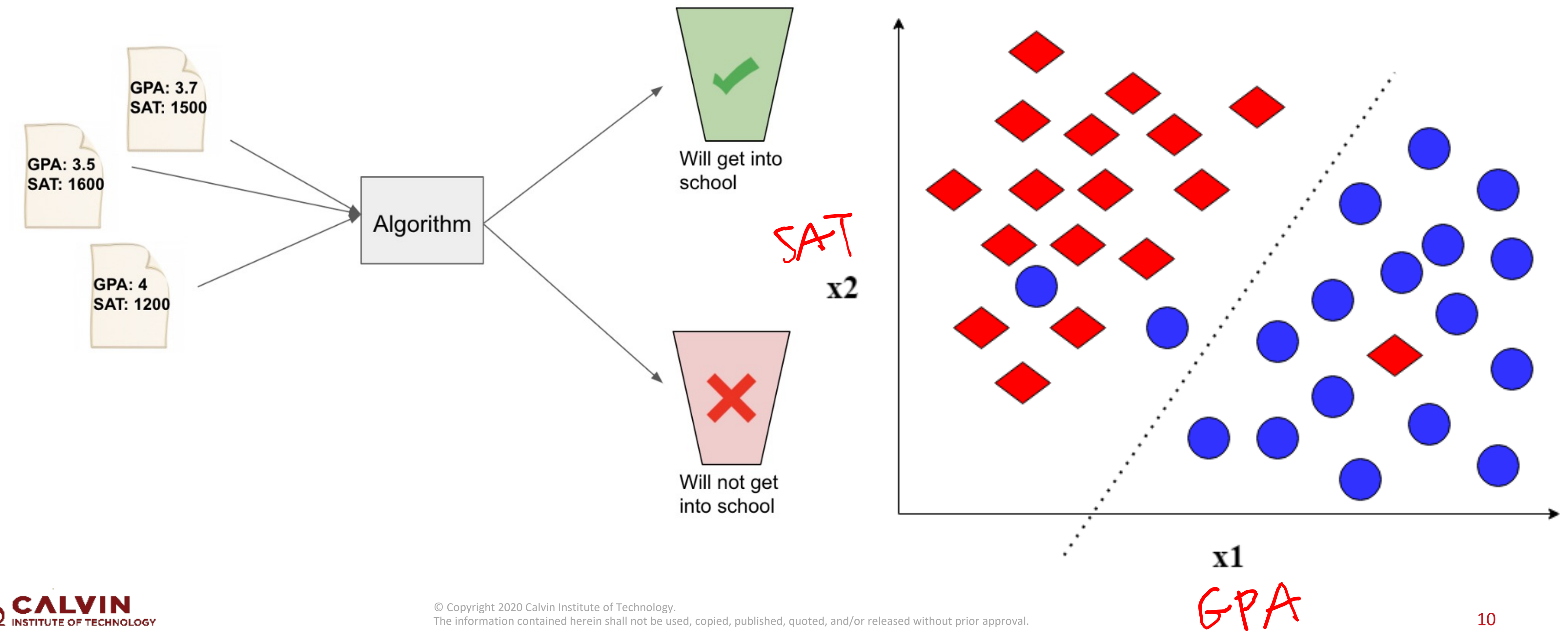
- Apakah jenis klasifikasi untuk memprediksi genre movie?

multilabel

Klasifikasi Biner

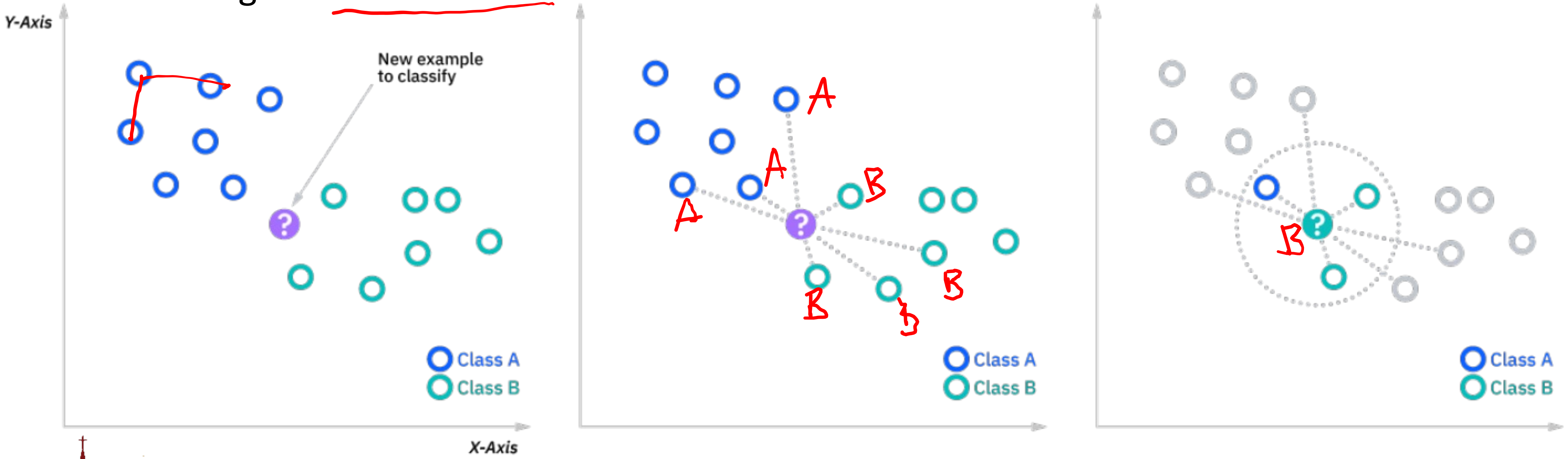
Klasifikasi Biner

- Model yang dibuat dapat membagi data menjadi 2 kelas



K-Nearest Neighbor

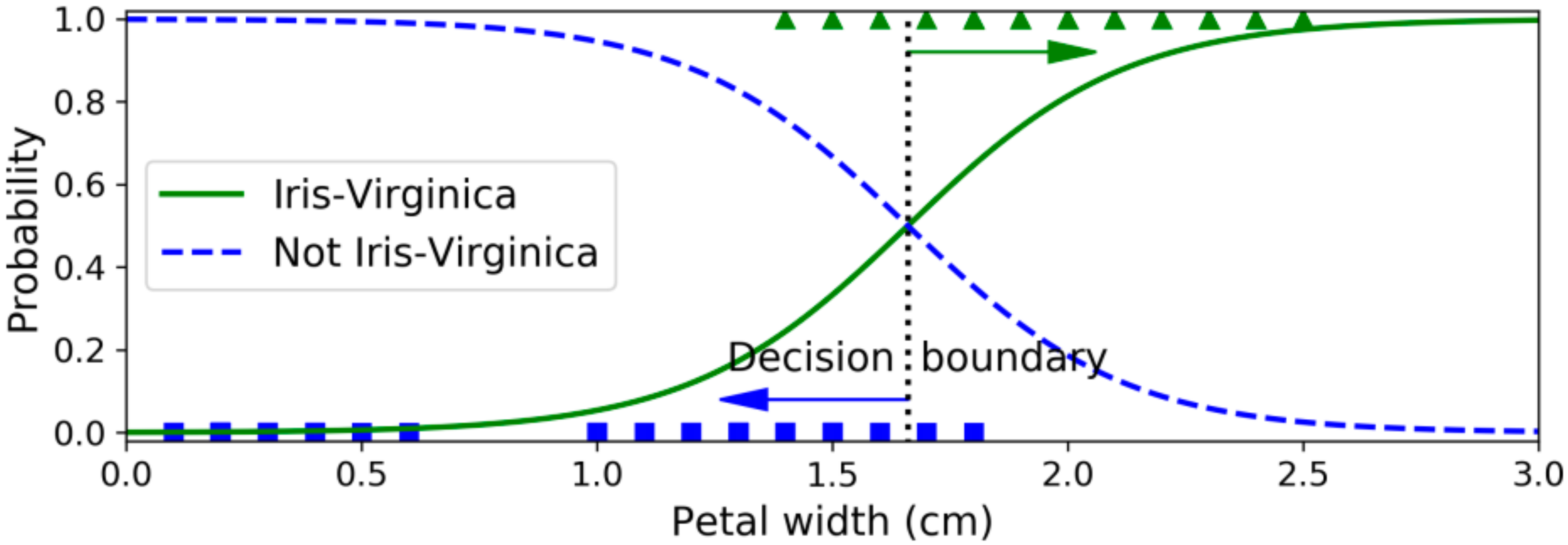
- $\boxed{\text{Lazy}}$ learning No learning
- Hyperparameter: k (k terlalu besar →underfitting, k terlalu kecil →overfitting)
- Kelebihan: mudah diimplementasikan
- Kekurangan: tidak scalable



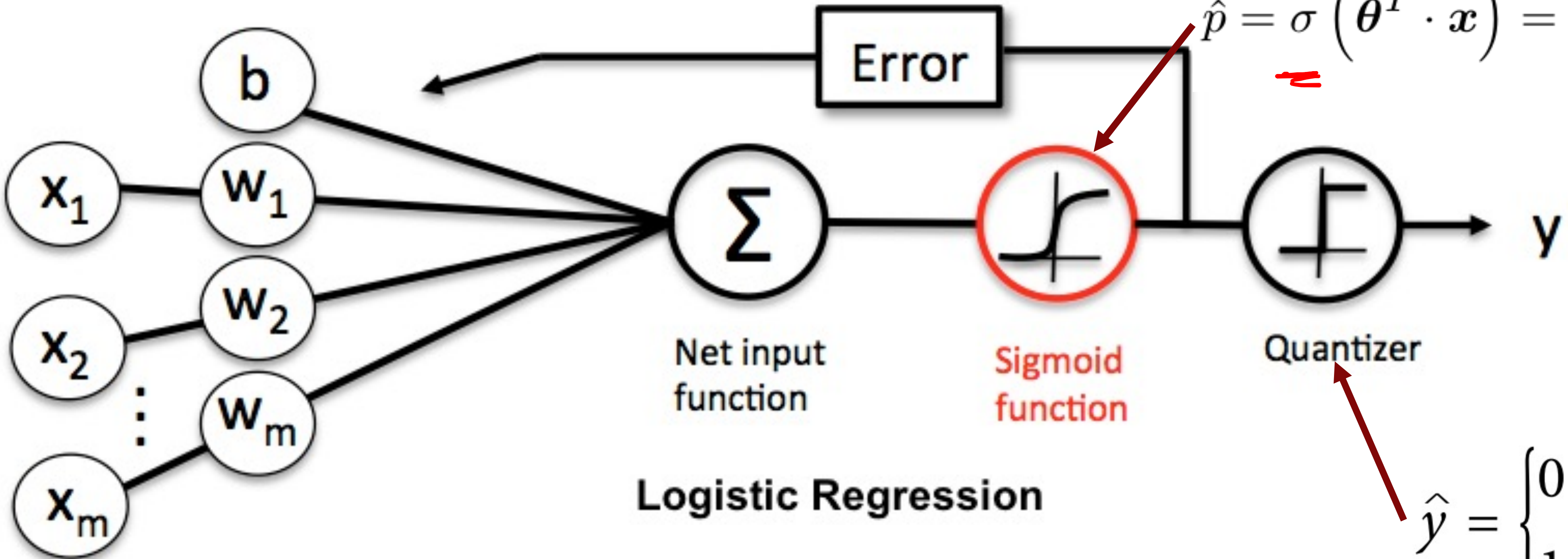
Regresi Logistik

$\theta_0 1 + \theta_1 x_1 + \theta_2 x_2 = \theta^T \cdot x$

$\sigma(\ ) = \frac{1}{1+e^{-(\ )}}$



$$\hat{p} = \sigma\left(\boldsymbol{\theta}^T \cdot \boldsymbol{x}\right) = \frac{1}{1+\exp(-\boldsymbol{\theta}^T \cdot \boldsymbol{x})}$$



$$\hat{y} = \begin{cases} 0 \text{ if } \hat{p} < 0.5 \\ 1 \text{ if } \hat{p} \geq 0.5 \end{cases}$$

Gradient Descent

$\sigma(\theta^T . x)$

- Cost function menggunakan cross entropy:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(\hat{p}^{(i)}\right) + \left(1 - y^{(i)}\right)\log\left(1 - \hat{p}^{(i)}\right)\right]$$

0,1

- Turunan parsial:

$$\frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\left(\sigma\left(\boldsymbol{\theta}^T \boldsymbol{x}^{(i)}\right) - y^{(i)}\right)x_j^{(i)}$$

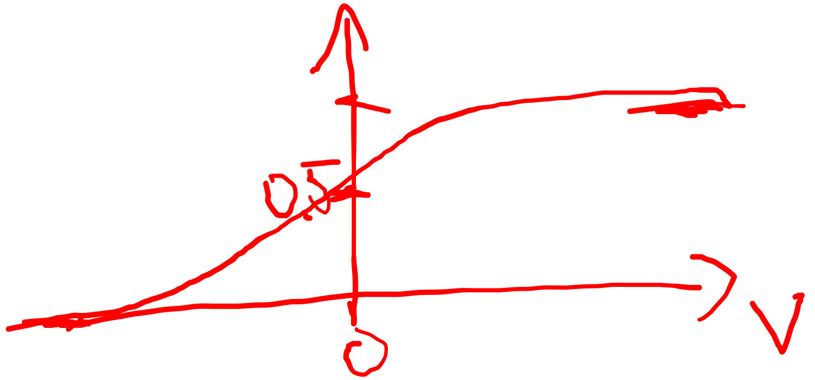
- Gradient Descent:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J\left(\theta_j\right) = \theta_j - \alpha \frac{1}{m}\sum_{i=1}^{m}\left(\sigma\left(\boldsymbol{\theta}^T \boldsymbol{x}^{(i)}\right) - y^{(i)}\right)x_j^{(i)}$$

$$\frac{1}{1+e^{-()}}$$

Uji Pemahaman

- Sebuah pembelajaran regresi logistik terhadap data barang rusak ($y = 1$ →rusak, $y = 0$ →tidak rusak) menghasilkan $\theta_0 = 0, \theta_1 = 1, \theta_2 = 2$. Apakah hasil prediksi dari:
  - $x_1 = 0, x_2 = 0$
  - $x_1 = 1, x_2 = 2$
  - $x_1 = 1, x_2 = -2$
- Apakah bentuk persamaan dari decision boundary nya?



$$h(x_1, x_2) = \sigma(\theta_0 1 + \theta_1 x_1 + \theta_2 x_2)$$

$$\sigma(0 + 1\cdot 0 + 2\cdot 0) = 0{,}5$$

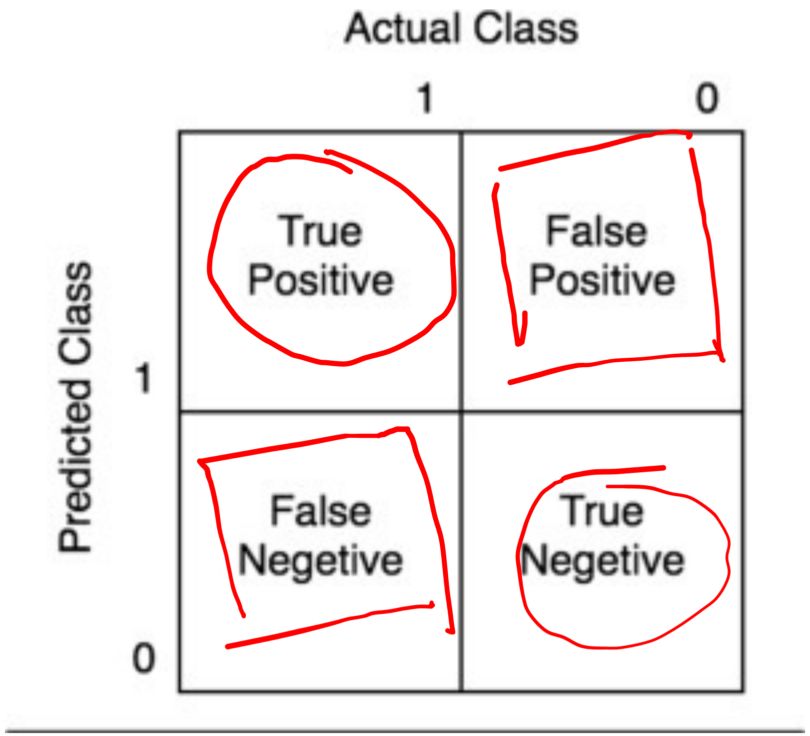
$$\sigma(0 + 1\cdot 1 + 2\cdot 2) = \sigma(5) \approx 1$$

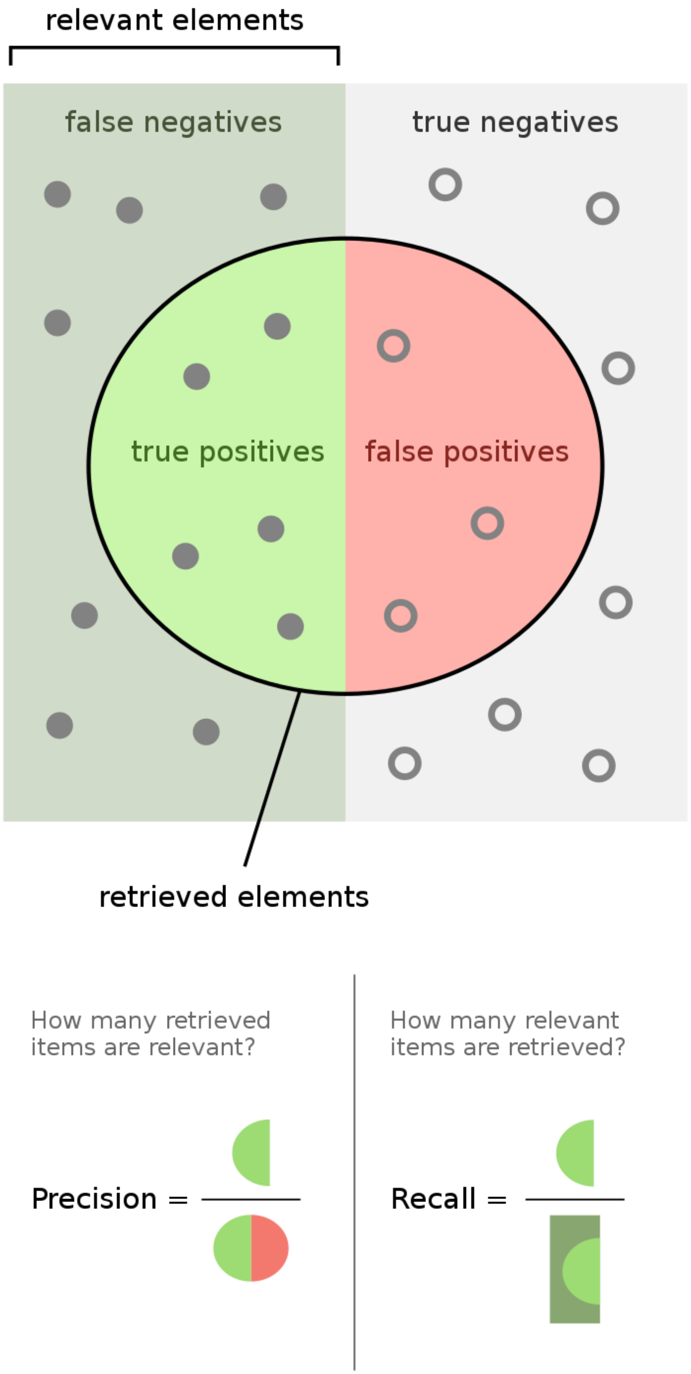
$$\sigma(0 + 1\cdot 1 + 2\cdot(-2)) = \sigma(-3) \approx 0$$

CALVIN INSTITUTE OF TECHNOLOGY

Metrik Performa

- Confusion matrix:
  - True positives
  - False positives
  - False negatives
  - True negatives





- $accuracy = \frac{TP+TN}{TP+FP+FN+TN}$
- $precision = \frac{TP}{TP+FP}$
- $recall = \frac{TP}{TP+FN}$
- $F1 = 2\frac{precision \times recall}{precision+recall}$

CALVIN INSTITUTE OF TECHNOLOGY

Uji Pemahaman

- Manakah yang lebih penting, menangkap false positive atau false negative?

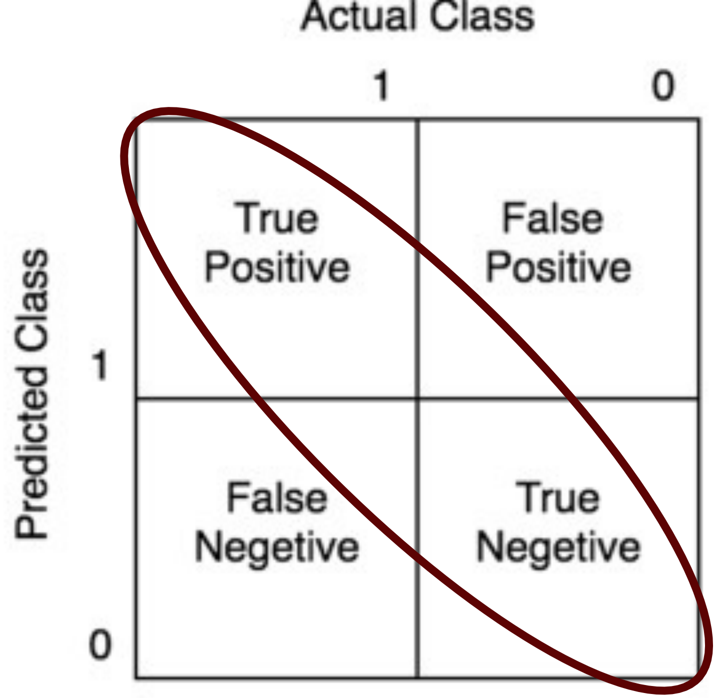
tergantung situasinya

Accuracy

0.01%

0.01%

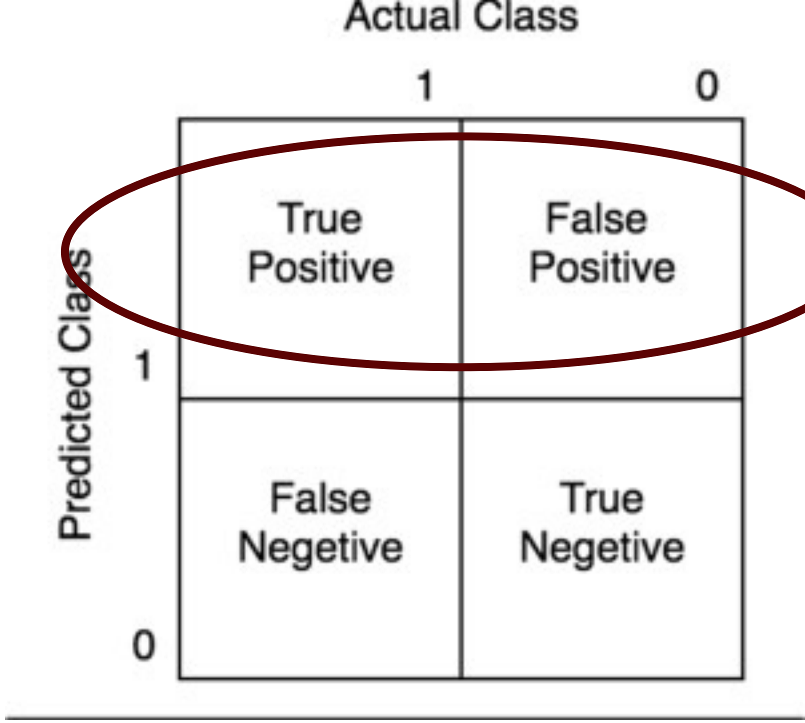
- $accuracy = \frac{TP+TN}{TP+FP+FN+TN}$
- Berapa % hasil yang diprediksi tepat



Well, here's the model: simply label every single person flying from a U.S. airport as "not a terrorist." Given the 800 million average passengers on U.S. flights per year and the 19 (confirmed) terrorists who boarded U.S. flights from 2000–2017, this model achieves an astounding accuracy of 99.9999999 percent! That might sound impressive, but I have a suspicion the Department of Homeland Security will not be calling anytime soon to buy this model. While this solution has nearly perfect accuracy, this problem is one in which accuracy is clearly not an adequate metric.

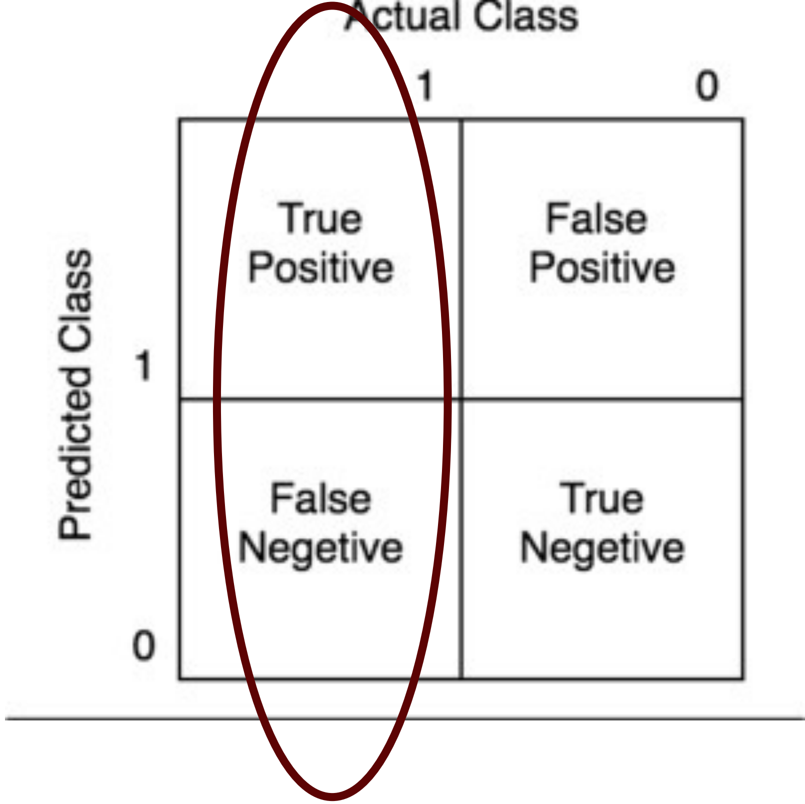
Precision

- $precision = \frac{TP}{TP+FP}$
- Berapa % kasus yang diprediksi positif benar-benar positif
- High precision diperlukan ketika harga yang harus dibayar untuk kasus false positive tinggi
- Contoh false positive: delete email yang bukan spam → lebih baik terima spam daripada delete email penting



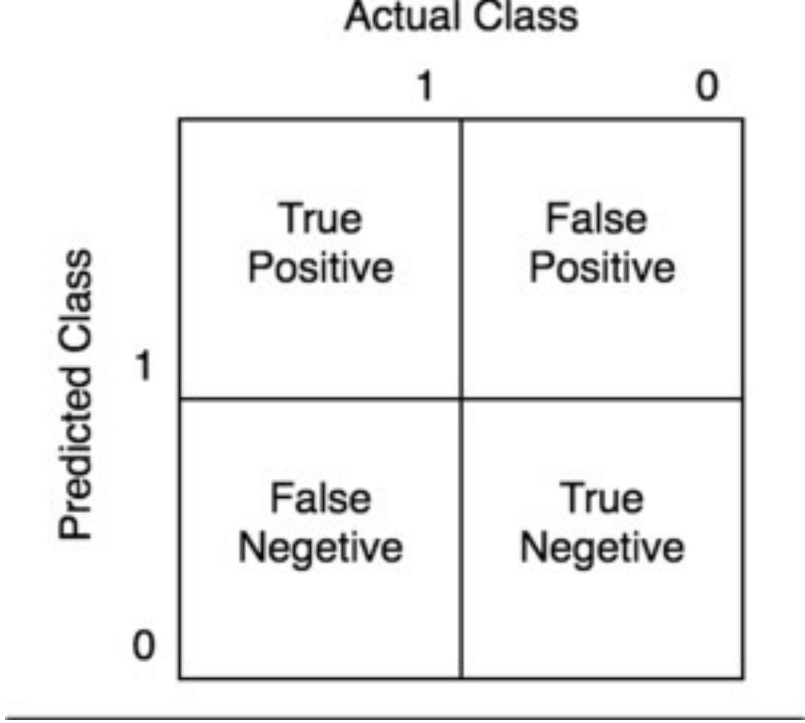
Recall

- $recall = \frac{TP}{TP+FN}$
- Berapa % kasus yang benar-benar positif dapat teridentifikasi dengan tepat
- High recall diperlukan Ketika harga yang harus dibayar untuk false negative tinggi
- Contoh false negative: tidak diprediksi ada tsunami padahal akan terjadi tsunami→lebih baik berjaga-jaga untuk evakuasi daripada memakan korban jiwa



F1

- $F1 = 2\frac{precision \times recall}{precision + recall}$
- Harmonic mean of precision and recall
- Balance antara precision dna recall
- High F1 diperlukan jika harga yang harus dibayar untuk false positive dan flase negative sama tinggi
- Misalnya quality control→terlalu kaku banyak barang bagus yang terbuang, terlalu longgar banyak barang rusak yang lolos

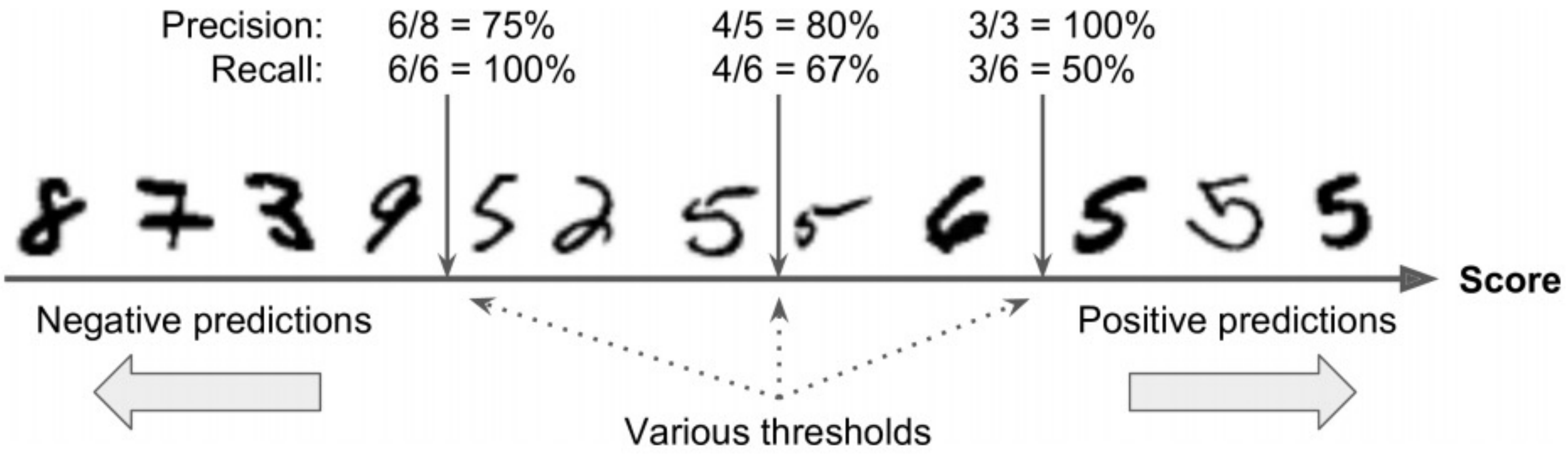


Studi kasus

- Klasifikasi apakah suatu box pasta gigi kosong atau tidak?
- Menggunakan x-ray memakan biaya mahal

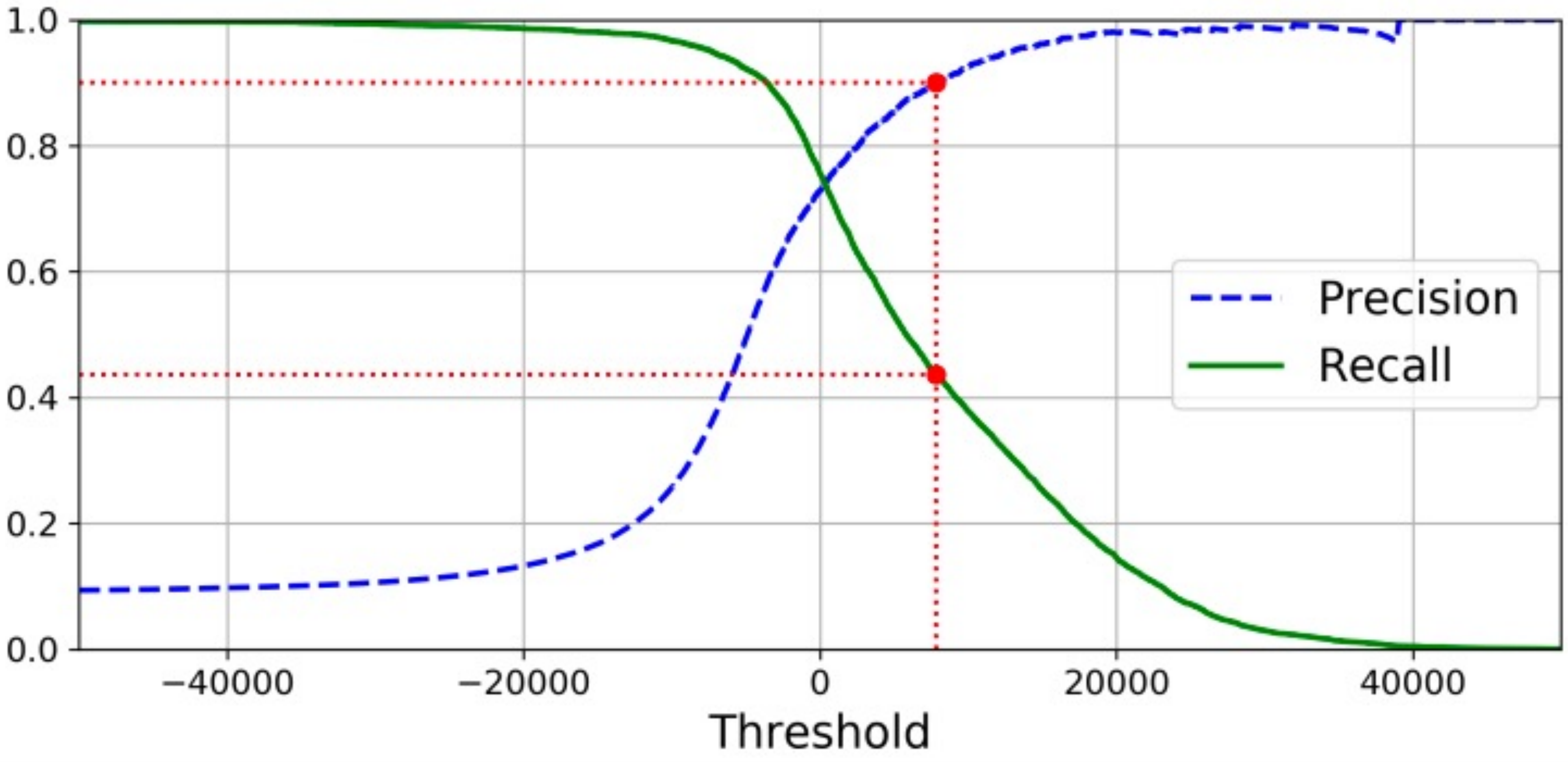
Precision-Recall Tradeoff

- Meningkatkan precision akan menurunkan recall dan sebaliknya
- Manakah boundary decision yang tepat untuk menghasilkan klasifikasi yang baik?
- Misal kita pilih 50% threshold →80% precision, 67% recall
- Misal kita pilih lebih tinggi →100% precision, 50% recall
- Misal kita pilih lebih rencah → 75% precision, 100% recall



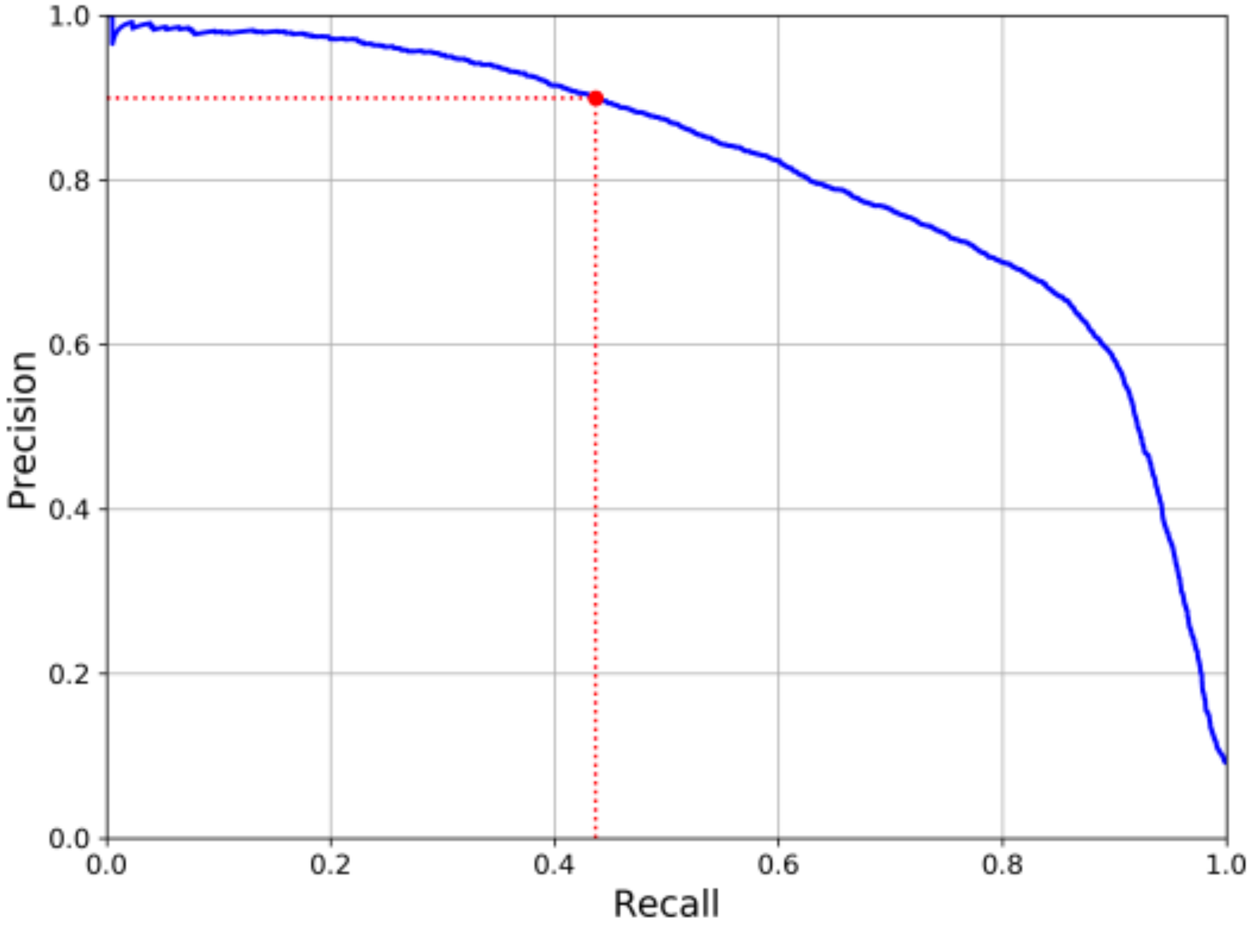
Precision-Recall Curve

- Dapatkan skor untuk semua data dalam training set, hitung precision dan recall → plot precision-recall curve
- Titik merah menunjukan 90% precision, 43% recall



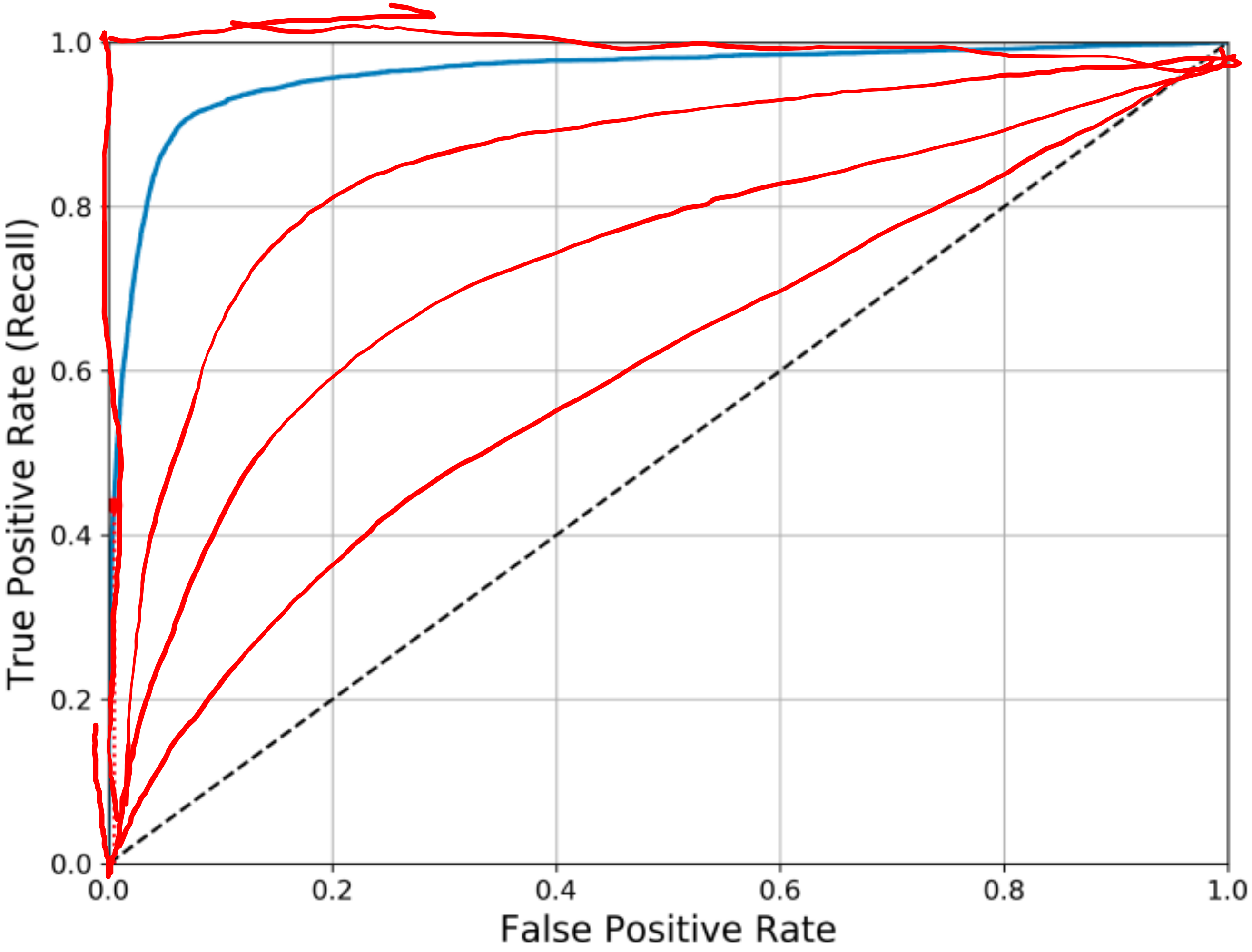
Precision vs recall curve

- Plot precision vs recall untuk berbagai threshold (titik merah: 90% precision, 43% recall)
- Kurva menurun tajam >80% recall (disarankan memilih threshold sebelum titik curam)



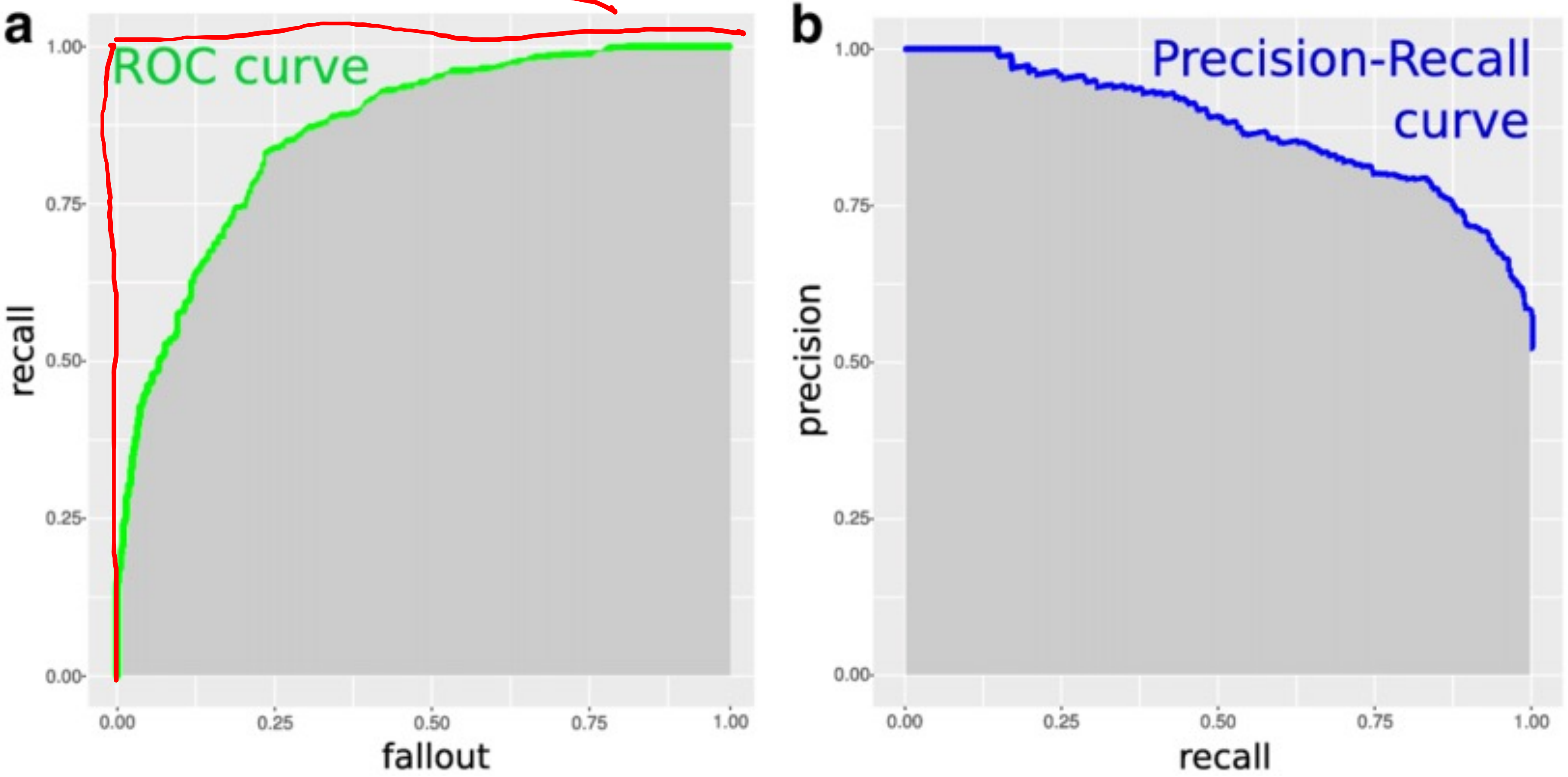
Receiver operating characteristics (ROC)

- Plot TPR vs FPR untuk berbagai nilai thresholds
- Sumbu vertikal: true positive rate
- Sumbu horizontal: false positive rate
- Garis putus-putus: 50% prediksi benar



Area Under Curve (AUC)

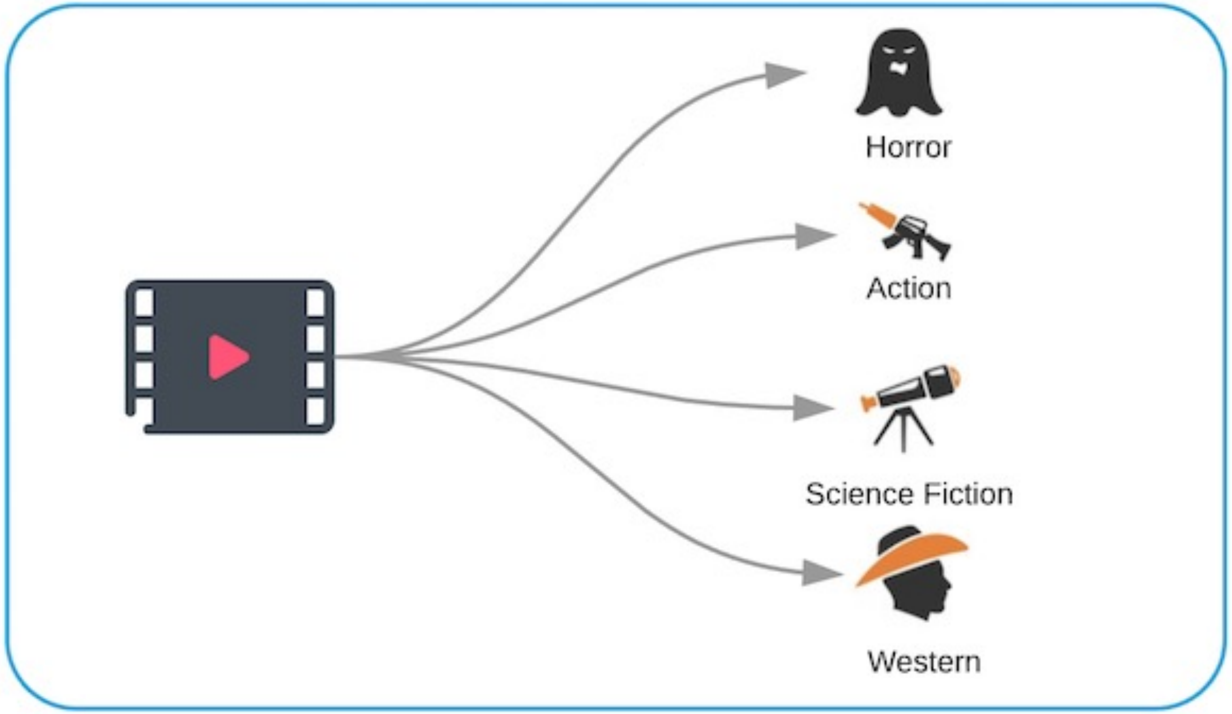
- Menghitung luas area dibawah suatu kurva
- Nilai AUC akan maksimal (AUC=1) jika kurva menyentuh pojok



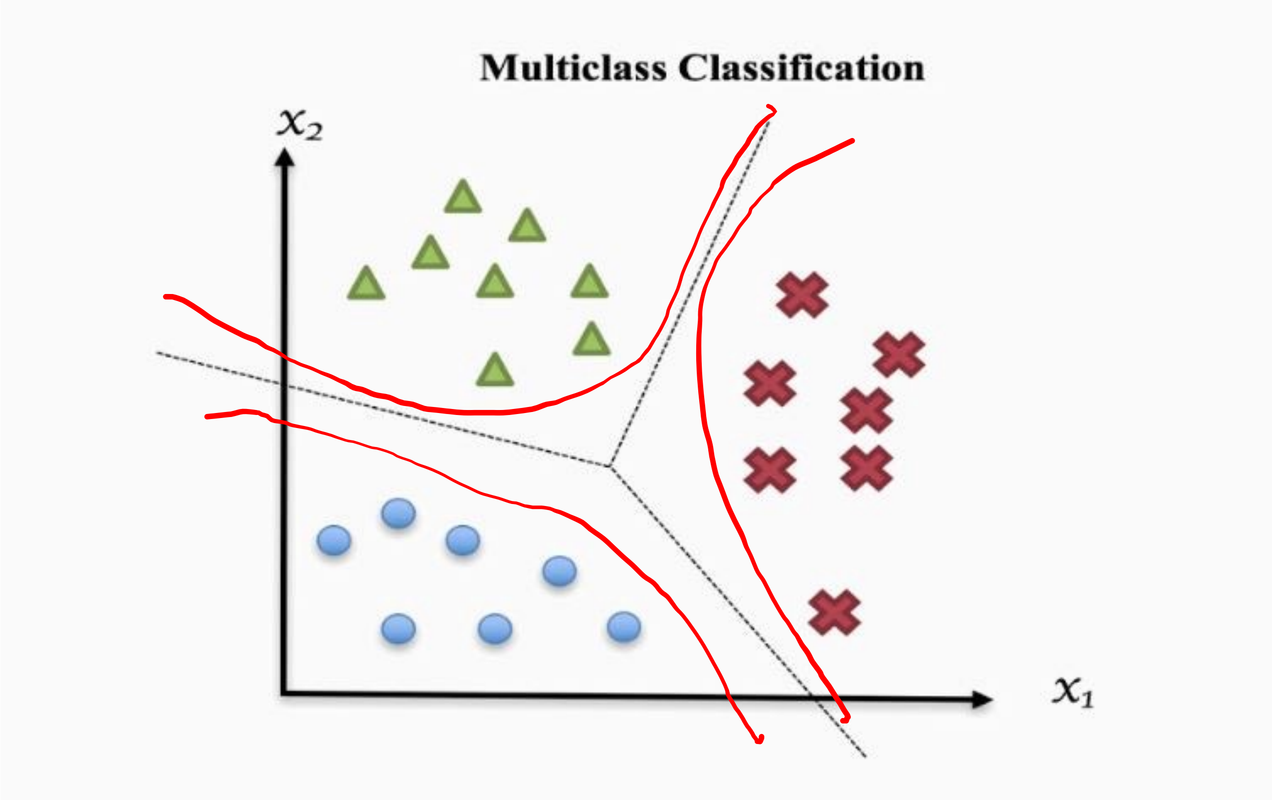
Klasifikasi Multi-Kelas

Klasifikasi Multi-Kelas

- Model yang dibuat dapat membagi data menjadi K kelas



Multi-class classification



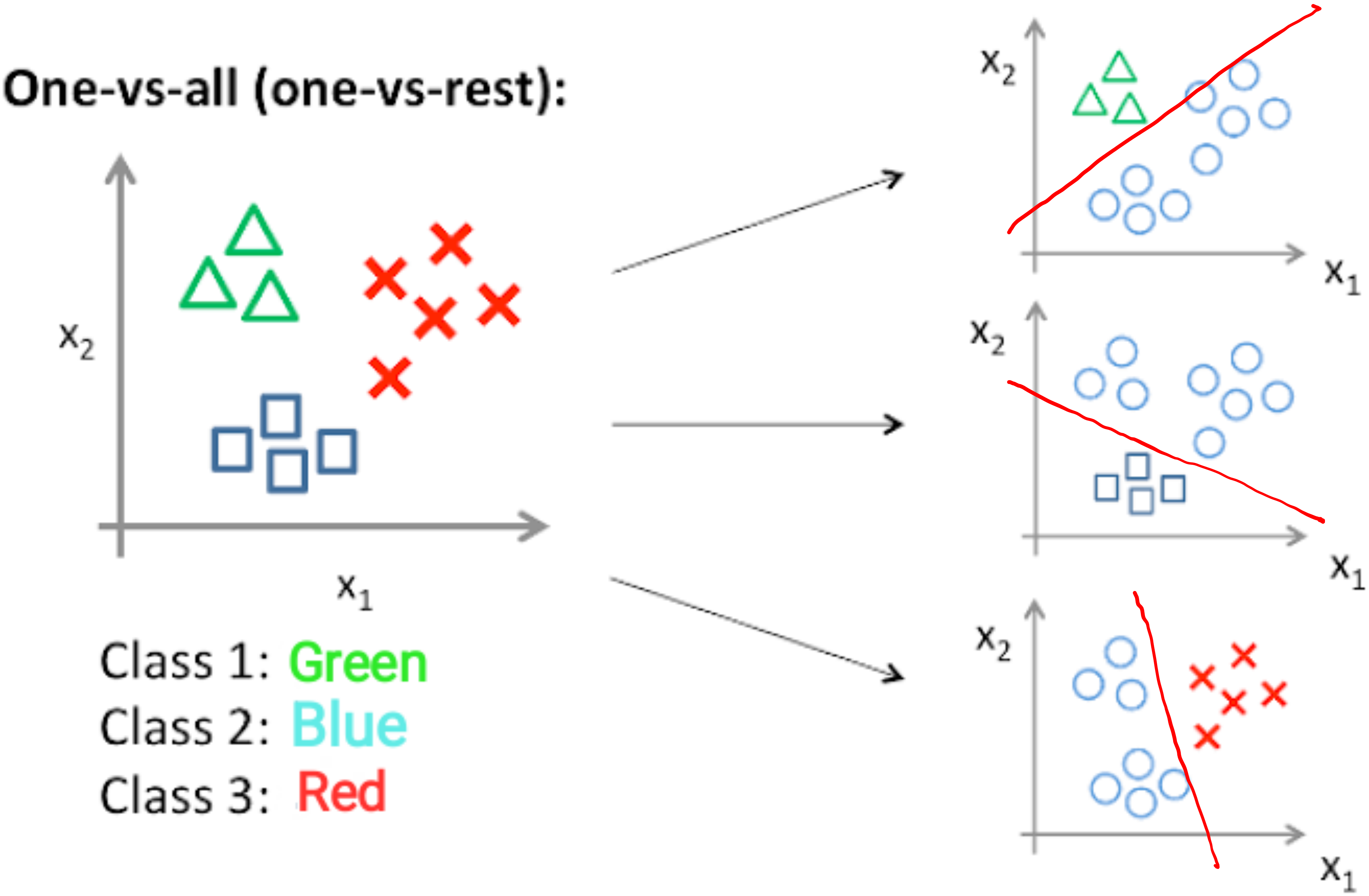
Pendekatan

A
B
C

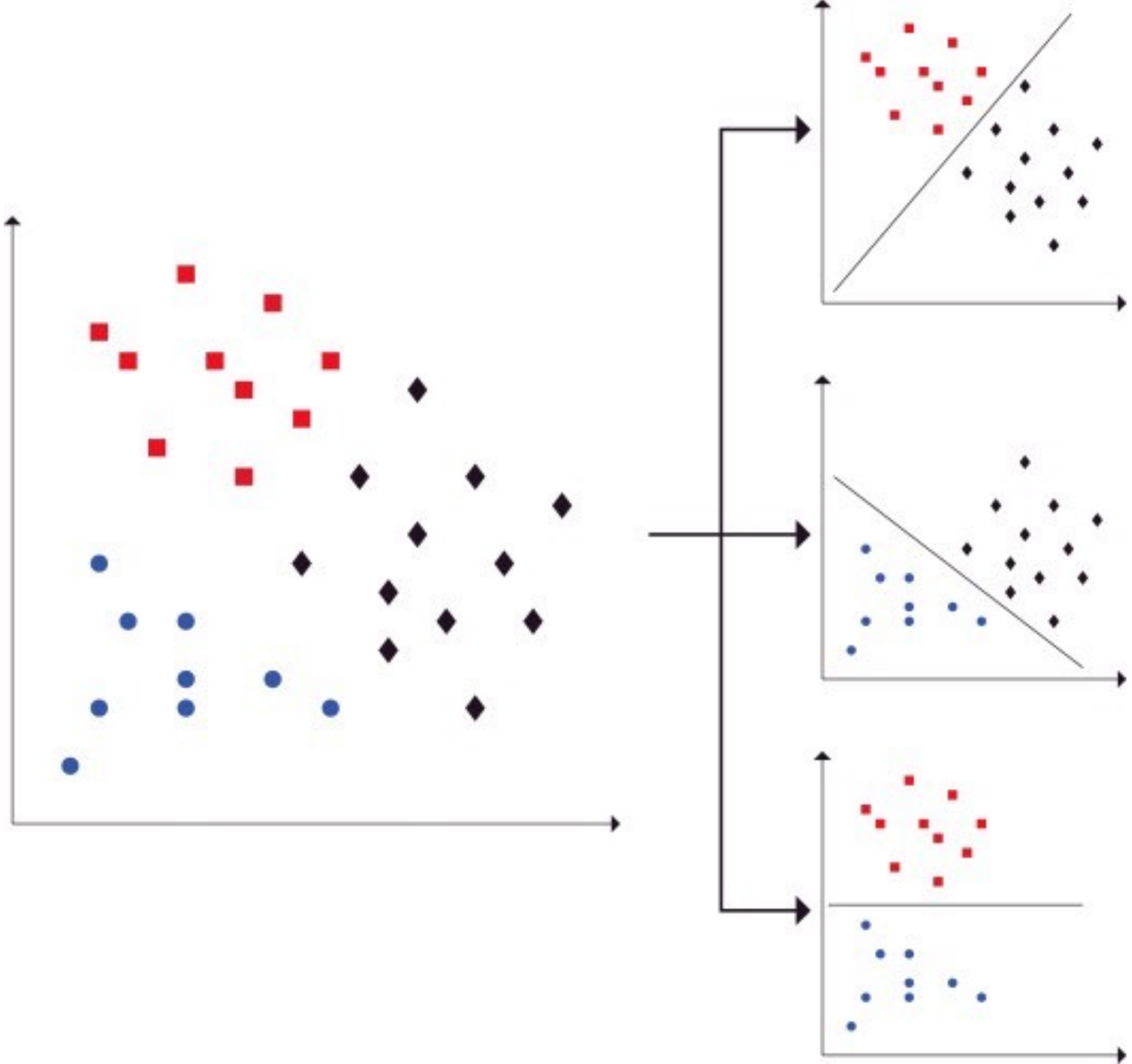
A vs BC
B vs AC
C vs AB

- Binary (one vs all): K-classifiers on large training sets (0-detectors, 1-detectors, …, K-detectors)→hasil: skor dari setiap classifier→ambil nilai tertinggi
- Binary (one vs one): $\frac{K(K-1)}{2}$-classifiers on small training sets (0 vs 1, 0 vs 2, …, K-1 vs K)→hasil: skor dari setiap duel→ambil kelas yang paling banyak menang duel
- Multiclass: 1-classifiers on entire training sets→hasil: skor dari setiap kelas→ambil nilai tertinggi

1 vs all


One-vs-all (one-vs-rest):
x2
x1
Class 1: Green
Class 2: Blue
Class 3: Red
x2
x1
x2
x1
x2
x1

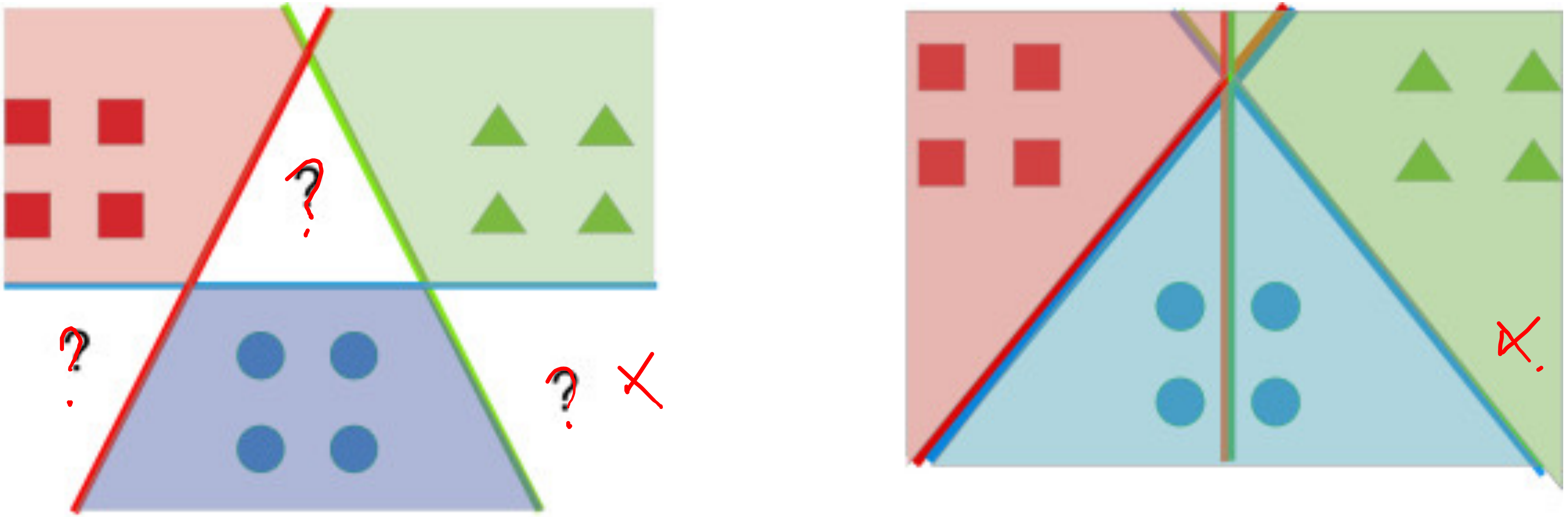
1 vs 1



OvA vs OvO

- OvA jauh lebih efisien, tapi dapat menghasilkan area ambigu yang sulit diprediksi

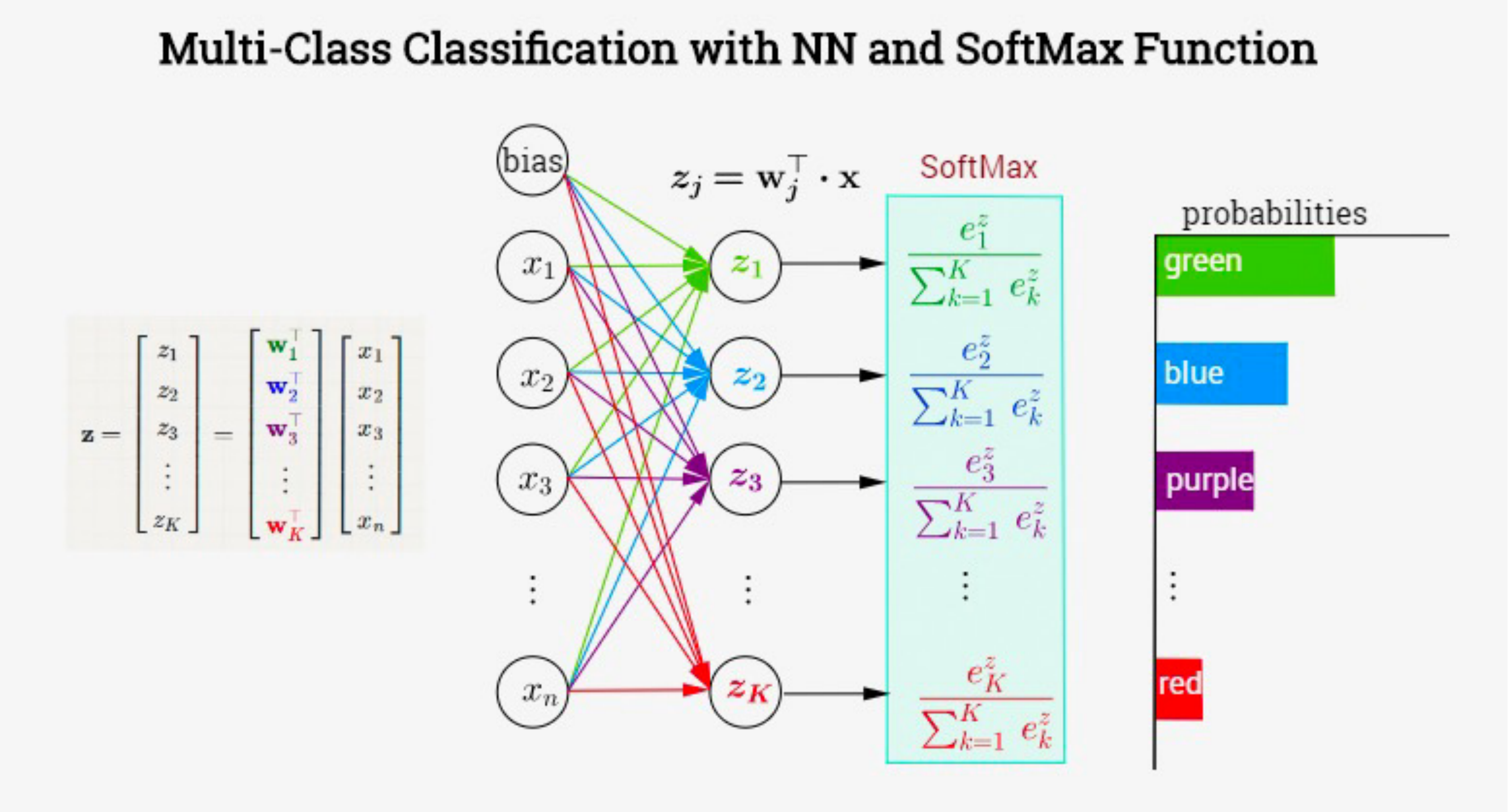
K $\qquad\qquad$ $K(K-1)/2$



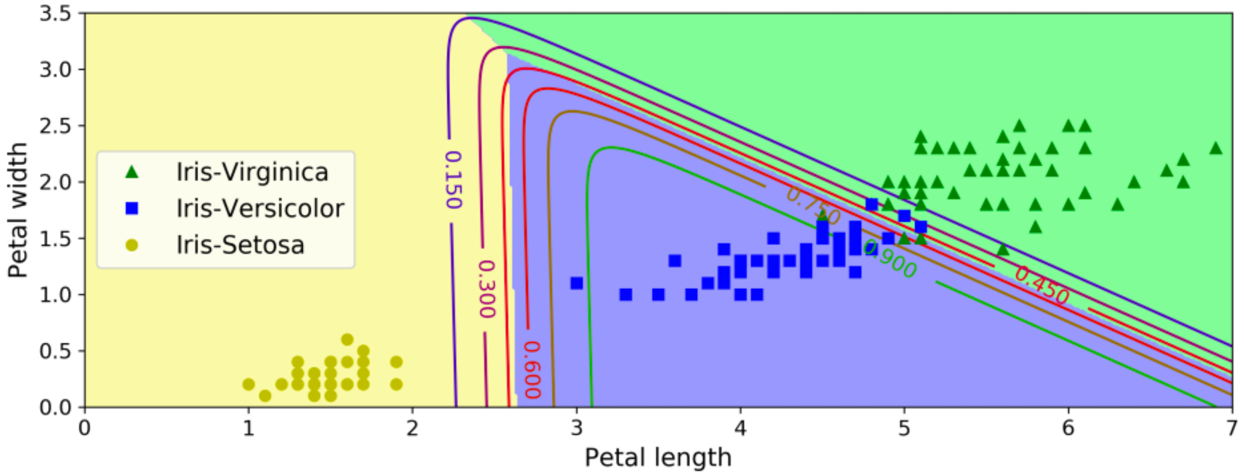
(a) Separation with OvA. (b) Separation with OvO.

Multiclass

- Arsitektur model dapat secara otomatis menghasilkan K kelas (tidak perlu melakukan binary classification berkali-kali)

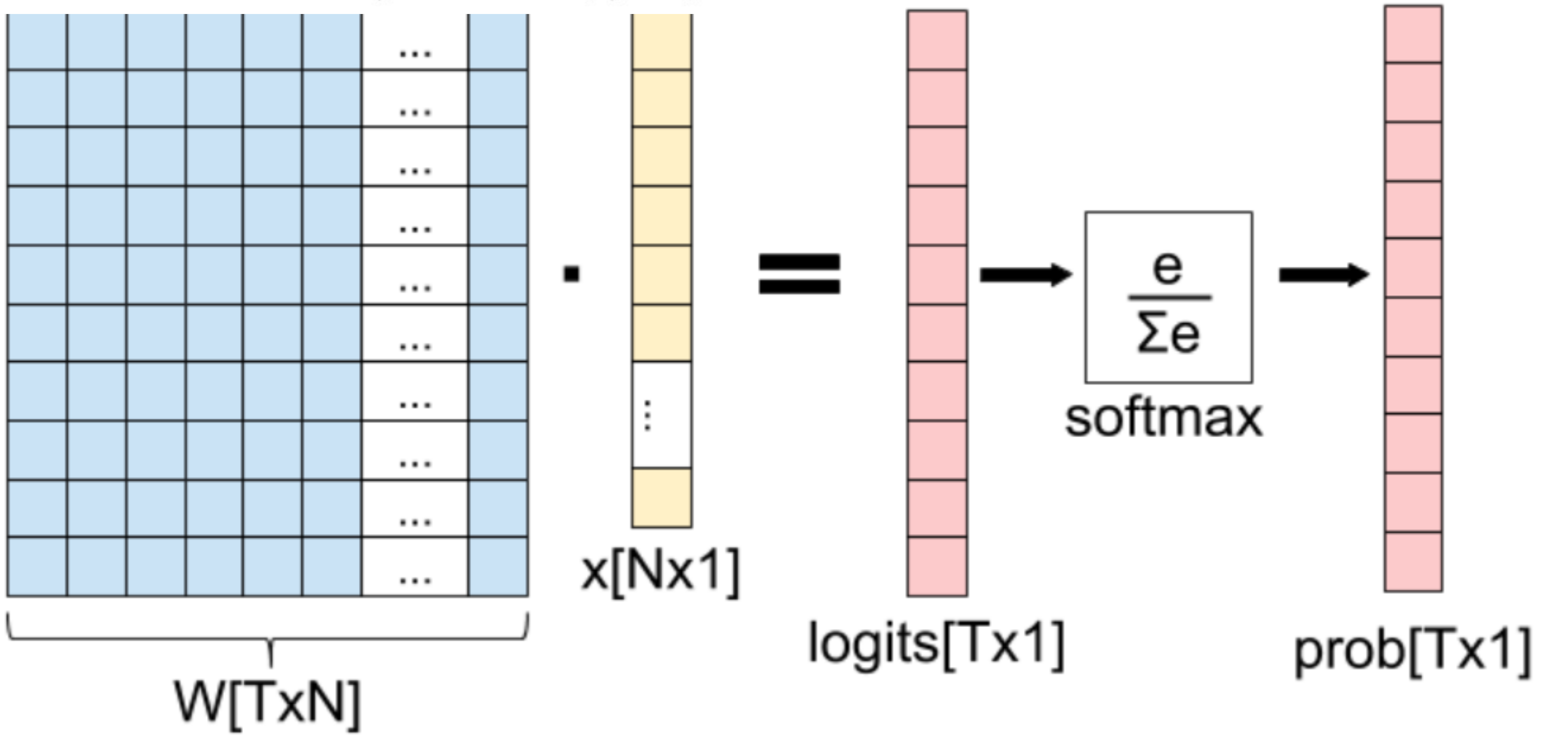


Regresi Softmax



- Peluang tiap kelas = softmax skor tiap kelas

$$\hat{p}_k = \sigma(\mathbf{s}(\mathbf{x}))_k = \frac{\exp\left(s_k(\mathbf{x})\right)}{\sum_{j=1}^{K} \exp\left(s_j(\mathbf{x})\right)}$$



Gradient Descent

- Cost function menggunakan cross entropy:

$$J(\boldsymbol{\Theta}) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log\left(\hat{p}_k^{(i)}\right)$$

- Turunan parsial:

$$\frac{\partial}{\partial \theta_{jk}} J(\boldsymbol{\Theta}) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{p}_k^{(i)} - y_k^{(i)}\right) x_j^{(i)}$$

- Gradient Descent:

$$\theta_{jk} \coloneqq \theta_{jk} - \alpha \frac{\partial}{\partial \theta_{jk}} J\left(\theta_{jk}\right) = \theta_{jk} - \alpha \frac{1}{m}\sum_{i=1}^{m}\left(\hat{p}_k^{(i)} - y_k^{(i)}\right) x_j^{(i)}$$

CALVIN INSTITUTE OF TECHNOLOGY

Uji Pemahaman

- Berapakah hasil penjumlahan dari semua prediksi kelas ($\sum_k^K \hat{p}_k$) pada regresi softmax?

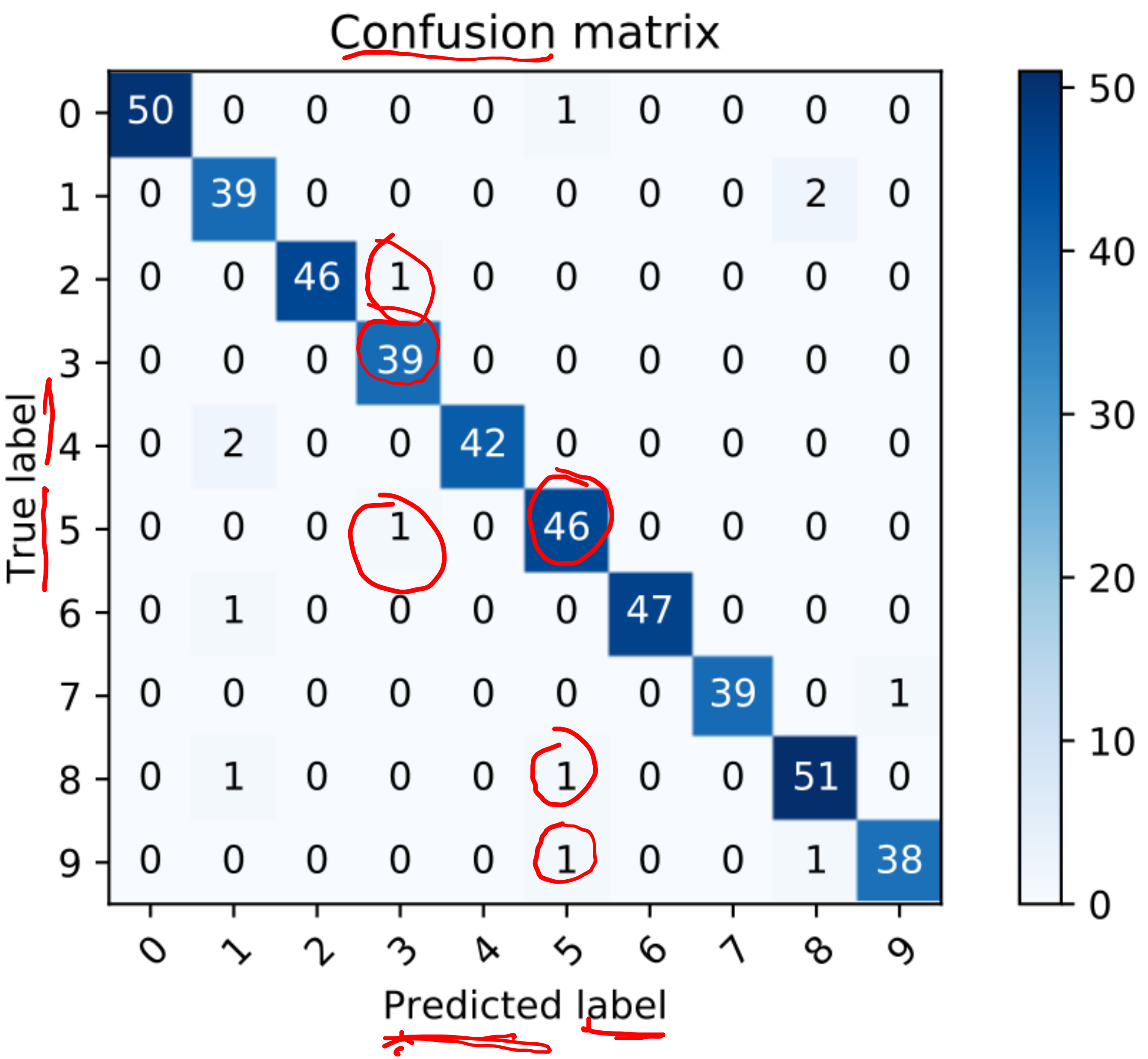
$$p_k = \frac{e^{-\theta_k^T \cdot X}}{\sum_k e^{-\theta_k^T \cdot X}}$$

$$\sum_{k=1}^{K} p_k = 1 = \sum_{k=1}^{K} \frac{\cancel{e^{-\theta_k^T \cdot X}}}{\cancel{\sum_{k=1}^{K} e^{-\theta_k^T \cdot X}}}$$

Metrik Performa

- Confusion matrix:
  - True positives
  - False positives
  - False negatives
  - True negatives

ACTUAL classification	PREDICTED classification			
Classes	a	b	c	d
a	TN	FP	TN	TN
b	FN	TP	FN	FN
c	TN	FP	TN	TN
d	TN	FP	TN	TN



Macro vs micro vs weighted

- https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f

```
              precision    recall  f1-score   support

   Aeroplane       0.67      0.67      0.67         3
        Boat       0.25      1.00      0.40         1
         Car       1.00      0.50      0.67         6

    accuracy                           0.60        10
   macro avg       0.64      0.72      0.58        10
weighted avg       0.82      0.60      0.64        10
```

Per-Class F1 scores

Average F1 scores

No	Actual	Predicted	Match
1	Airplane	Airplane	✓
2	Car	Boat	X
3	Car	Car	✓
4	Car	Car	✓
5	Car	Boat	X
6	Airplane	Boat	X
7	Boat	Boat	✓
8	Car	Airplane	X
9	Airplane	Airplane	✓
10	Car	Car	✓

Metrik tiap kelas

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Precision	Recall	F1 Score
Airplane	2	1	1	0.67	0.67	2 * (0.67 * 0.67) / (0.67 + 0.67) = **0.67**
Boat	1	3	0	0.25	1.00	2*(0.25 * 1.00) / (0.25 + 1.00) = **0.40**
Car	3	0	3	1.00	0.50	2 * (1.00 * 0.50) / (1.00 + 0.50) = **0.67**

$$\frac{0.67 + 0.25 + 1}{3} = \frac{1.92}{3} = 0.64$$

$$\frac{6}{6+4} = 0.6$$

$$0.3 \times 0.67 + 0.1 \times 0.25 + 0.6 \times 1 = 0.826$$

Micro-Averaged

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Micro-Averaged Values
Airplane	2	1	1	Precision $= \frac{6}{6+4} = \mathbf{0.60}$
Boat	1	3	0	Recall $= \frac{6}{6+4} = \mathbf{0.60}$
Car	3	0	3	
TOTAL	**6**	**4**	**4**	F1 Score $= \frac{6}{6+\frac{1}{2}(4+4)} = \mathbf{0.60}$

Macro vs micro vs weighted

- Contoh: F1

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Micro-Averaged F1 Score
Airplane	2	1	1	$\frac{TP}{TP+\frac{1}{2}(FP+FN)} = \frac{6}{6+\frac{1}{2}(4+4)}$ = **0.60**
Boat	1	3	0	
Car	3	0	3	
TOTAL	**6**	**4**	**4**	

Label	Per-Class F1 Score	Macro-Averaged F1 Score
Airplane	0.67	$\frac{0.67 + 0.40 + 0.67}{3}$ = **0.58**
Boat	0.40	
Car	0.67	

Label	Per-Class F1 Score	Support	Support Proportion	Weighted Average F1 Score
Airplane	0.67	3	0.3	$(0.67 * 0.3) + (0.40 * 0.1) + (0.67 * 0.6)$ = **0.64**
Boat	0.40	1	0.1	
Car	0.67	6	0.6	
Total	-	10	1.0	

Uji Pemahaman

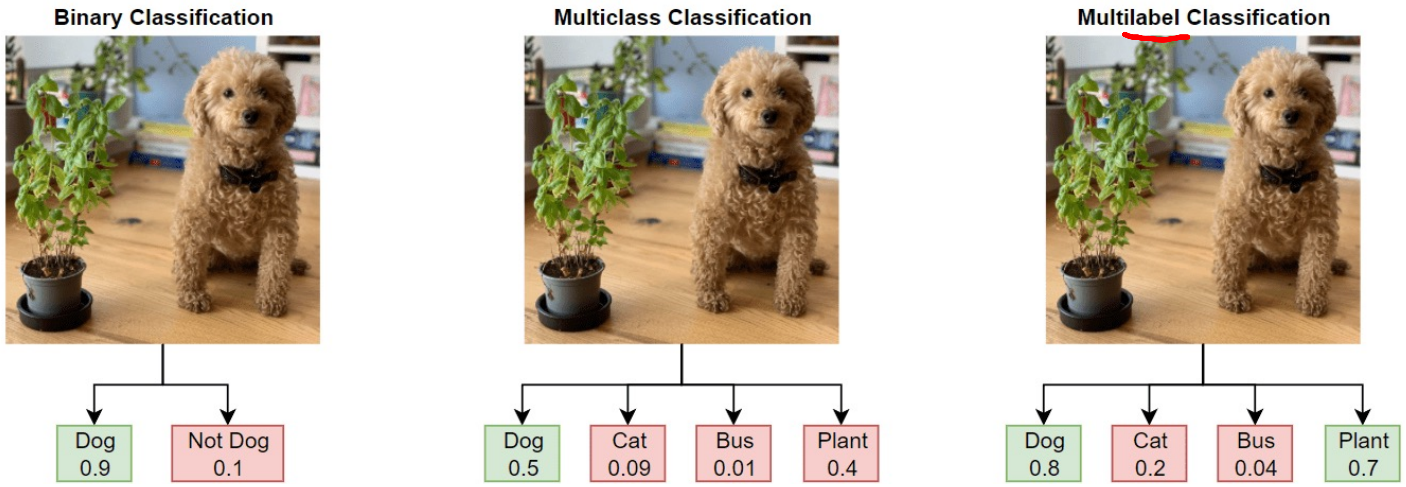
- Bagaimana cara menghitung accuracy, precision, recall, F1 untuk kasus weighted, macro, dan micro averaged

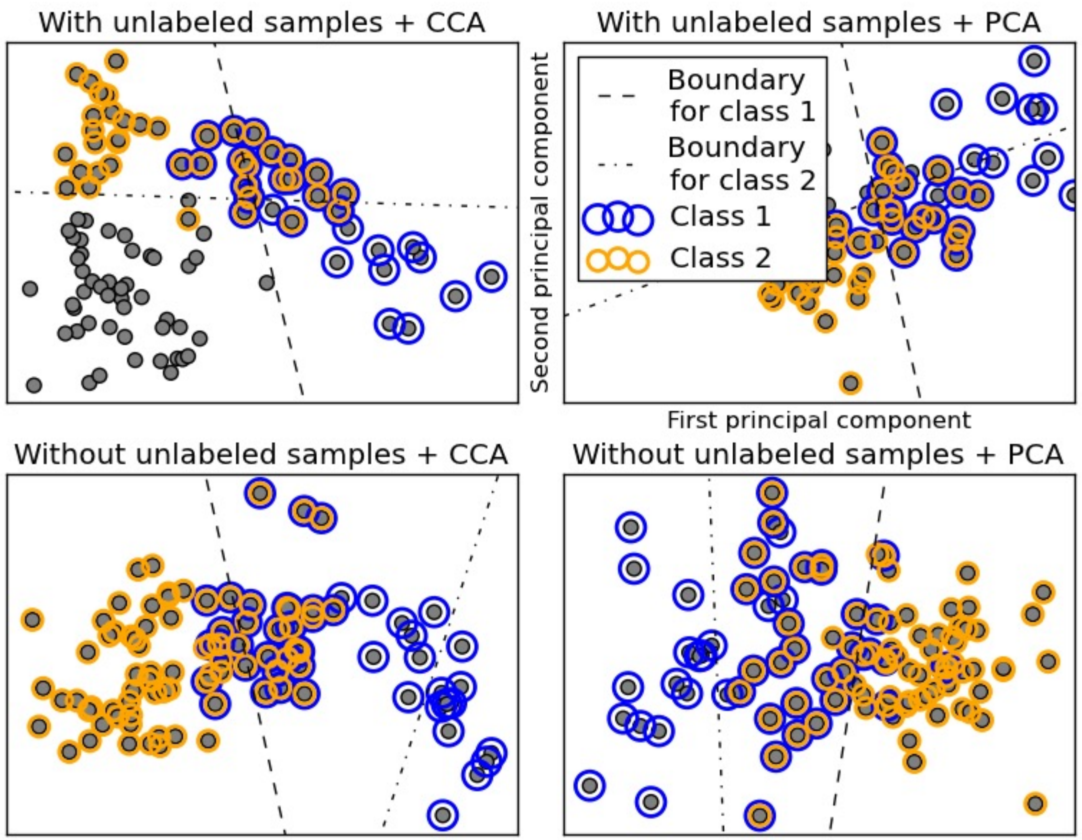
	Acc	Prec	Rec	F1
macro	0.6	0.64		0,6
micro	0.6	0.6		0.58
weighted	0.6	0.826		0,64

Klasifikasi Multi-Label

Klasifikasi Multi-Label

- Model yang dibuat dapat melabeli data dengan beberapa K kelas





Klasifikasi Multi-Label

- https://www.geeksforgeeks.org/an-introduction-to-multilabel-classification/

TEXT	SERVICE	FOOD	ANECDOTES	PRICE	AMBIENCE
but the staff was so horrible to us	1	0	0	0	0
to be completely fair the only redeeming facto...	0	1	1	0	0
the food is uniformly exceptional with a very ...	0	1	0	0	0
where gabriela personaly greets you and recomm...	1	0	0	0	0
for those that go once and dont enjoy it all i...	0	0	1	0	0

Accuracy

data

True labels

Prediksi model

Predicted labels

TEXT	True labels					Predicted labels				
	SERVICE	FOOD	ANECDOTES	PRICE	AMBIENCE	SERVICE	FOOD	ANECDOTES	PRICE	AMBIENCE
but the staff was so horrible to us	1	0	0	0	0	0	1	0	0	0
to be completely fair the only redeeming facto...	0	1	1	0	0	1	1	0	0	0
the food is uniformly exceptional with a very ...	0	1	0	0	0	0	0	0	1	0
where gabriela personaly greets you and recomm...	1	0	0	0	0	1	0	0	0	0
for those that go once and dont enjoy it all i...	0	0	1	0	0	1	0	0	0	0

Total number of predictions (TNP) = ~~4~~ 5

Total number of correct predictions (TNCP) = 1

1/5 = 0.2

Accuracy = TNCP/TNP = ~~¼~~ = ~~0.25~~

Hamming loss

5

TEXT	True labels					Predicted labels				
	SERVICE	FOOD	ANECDOTES	PRICE	AMBIENCE	SERVICE	FOOD	ANECDOTES	PRICE	AMBIENCE
but the staff was so horrible to us	1	0	0	0	0	0	1	0	0	0
to be completely fair the only redeeming facto...	0	1	1	0	0	1	1	0	0	0
the food is uniformly exceptional with a very ...	0	1	0	0	0	0	0	0	1	0
where gabriela personaly greets you and recomm...	1	0	0	0	0	1	0	0	0	0
for those that go once and dont enjoy it all i...	0	0	1	0	0	1	0	0	0	0

5

Total number of predictions (TNP) = 25 = 5×5

Total number of incorrect predictions (TNIP) = 8

Hamming loss

~~Accuracy~~ = TNIP/TNP = 8/25 = 0.32

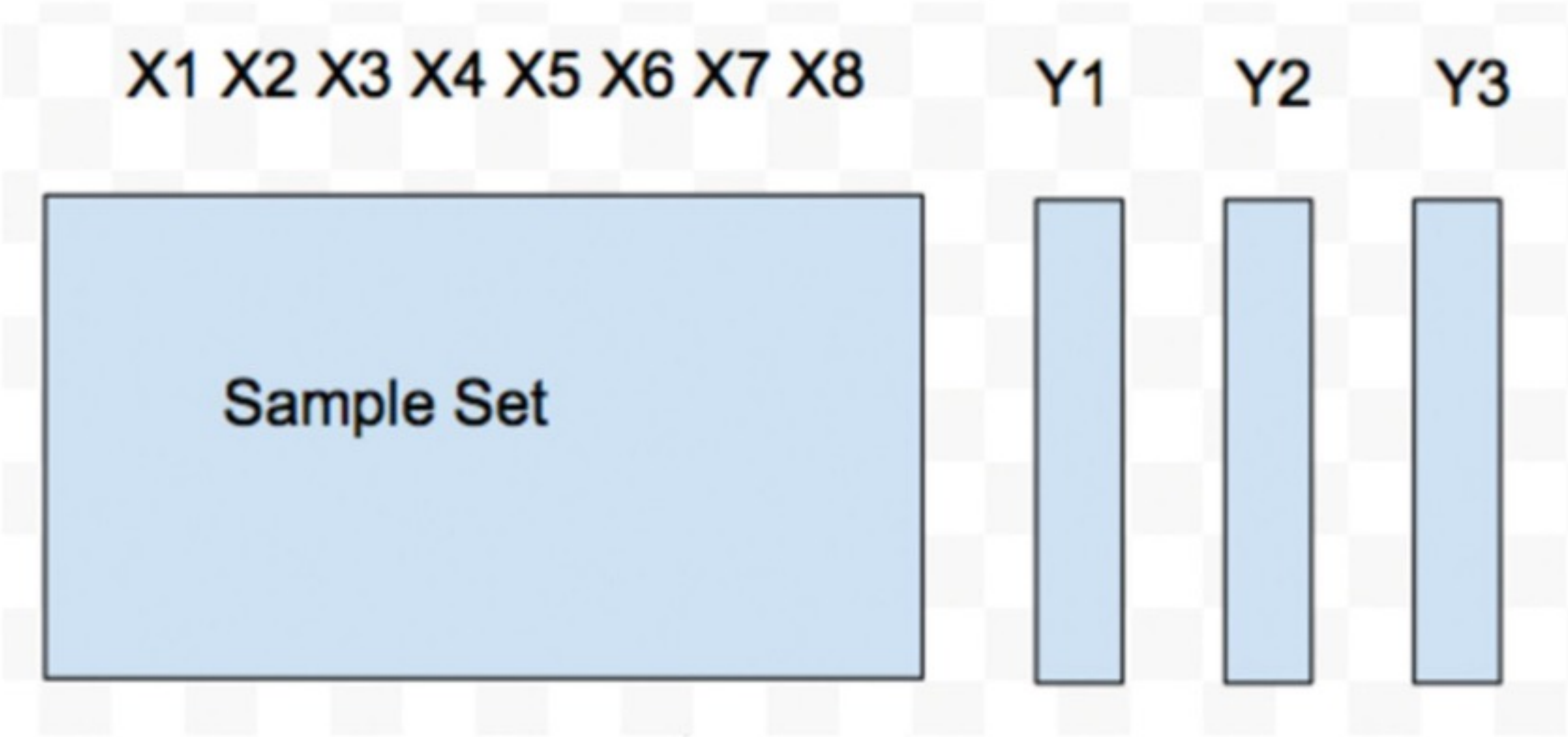
Klasifikasi Multi-Output

- Model yang dibuat dapat melakukan beberapa klasifikasi sekaligus
- Contoh dibawah ini adalah angka 4 dengan dan tanpa noise
- Model dapat membedakan angka tanpa dan dengan noise maupun angka berapa dari 10 pilihan angka

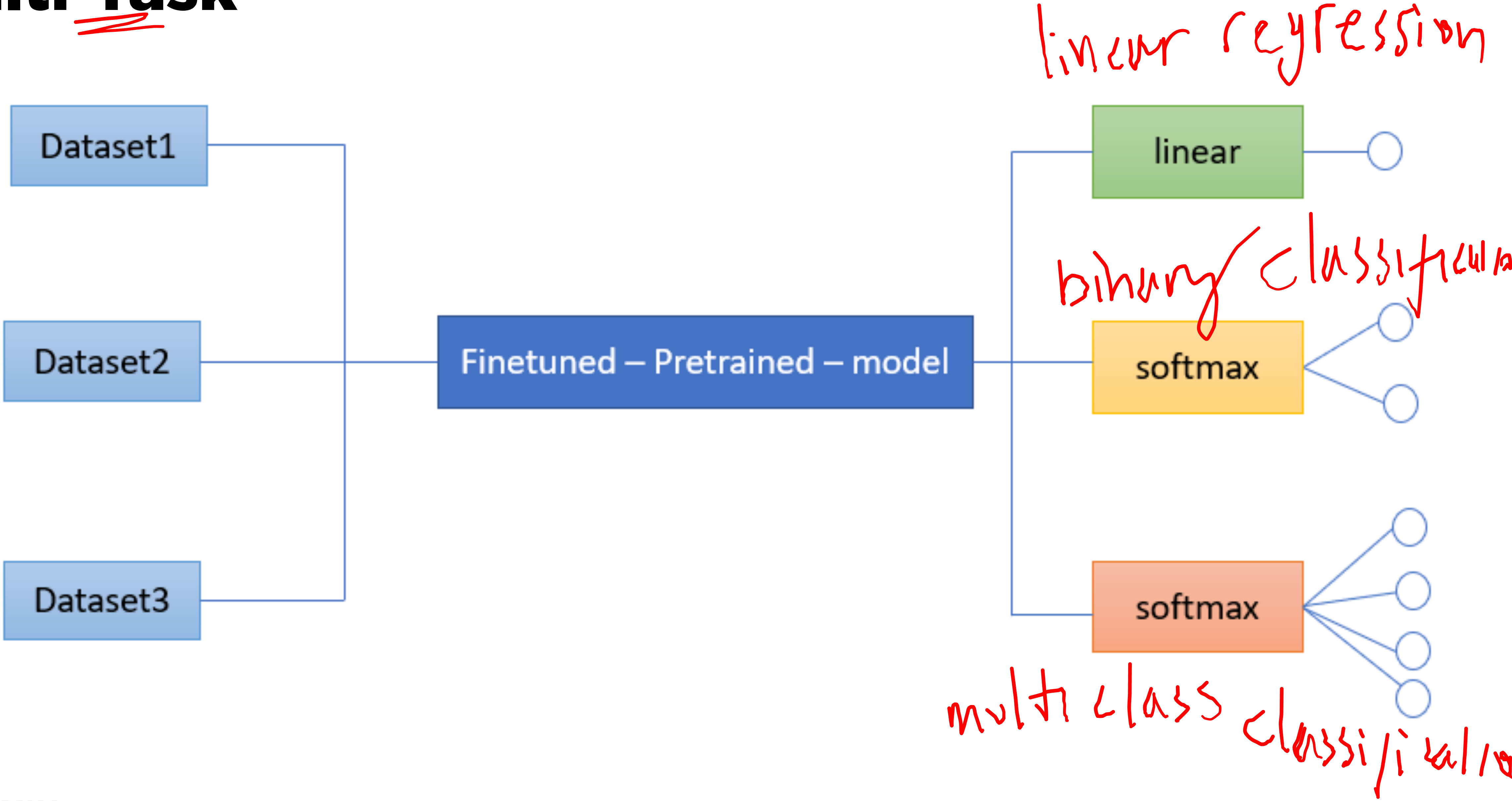
multiclass

binary

Regresi Multi-Output

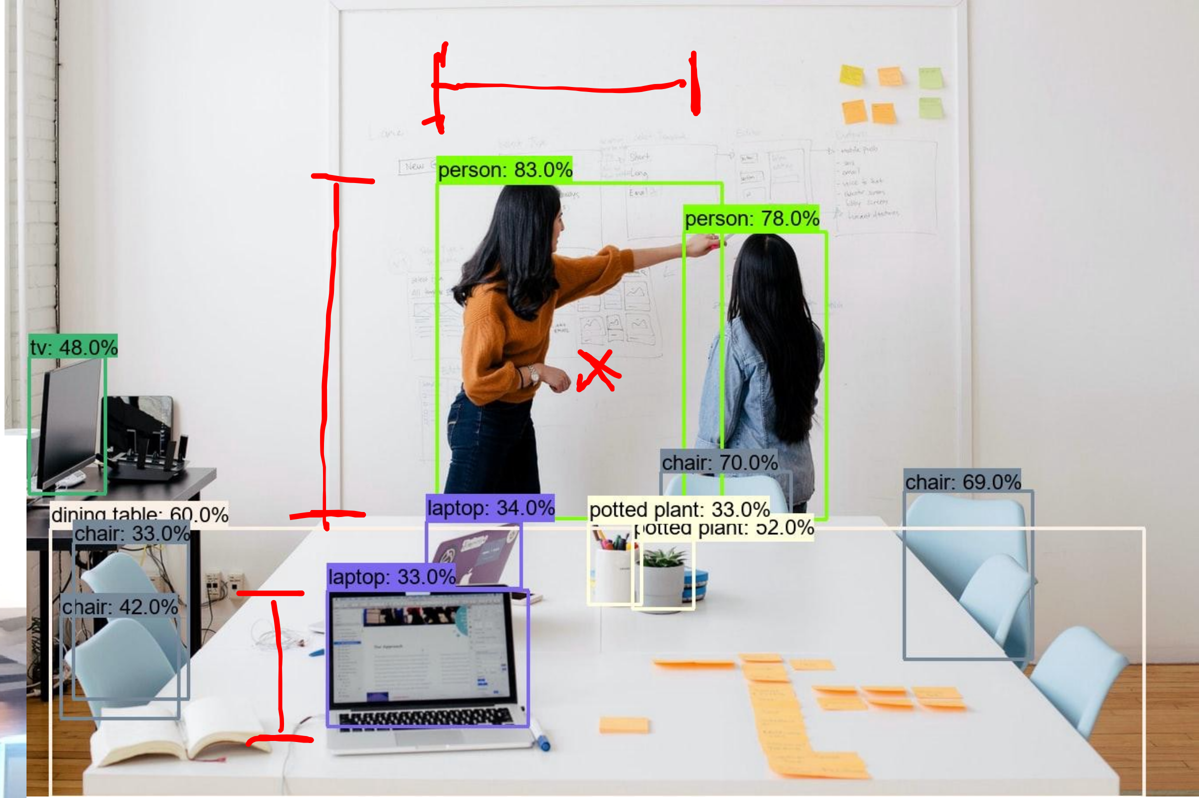


Multi-Task


linear regression
Dataset1
Dataset2
Dataset3
Finetuned – Pretrained – model
linear
binary classification
softmax
softmax
multi class classification

Contoh Multi-Task

- Contoh: deteksi objek, segmentasi
- Terdapat klasifikasi dan regresi



1. Klasifikasi objek
2. x, y pusat objek
3. $\Delta X, \Delta Y$ kotak
4. Segmentasi Objek

Contoh Cost Function YOLO

titik pusat objek prediksi

Regression loss

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

ukuran kotak prediksi

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

Confidence loss

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

objek prediksi

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

Classification loss

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

kelas objek prediksi

CALVIN INSTITUTE OF TECHNOLOGY

Tuhan Memberkati

God's People for God's Glory
CALVIN
INSTITUTE OF TECHNOLOGY