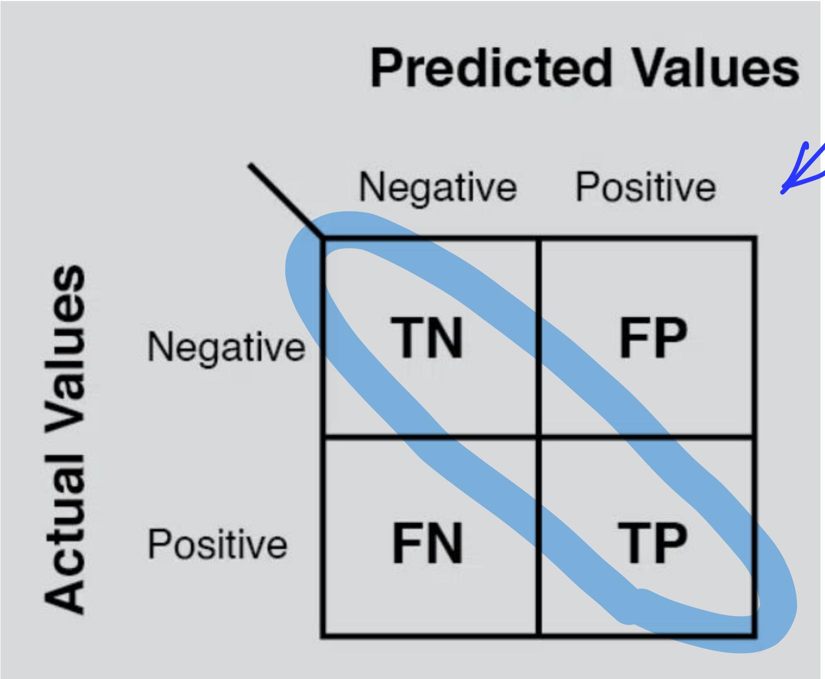
Ensemble Learning

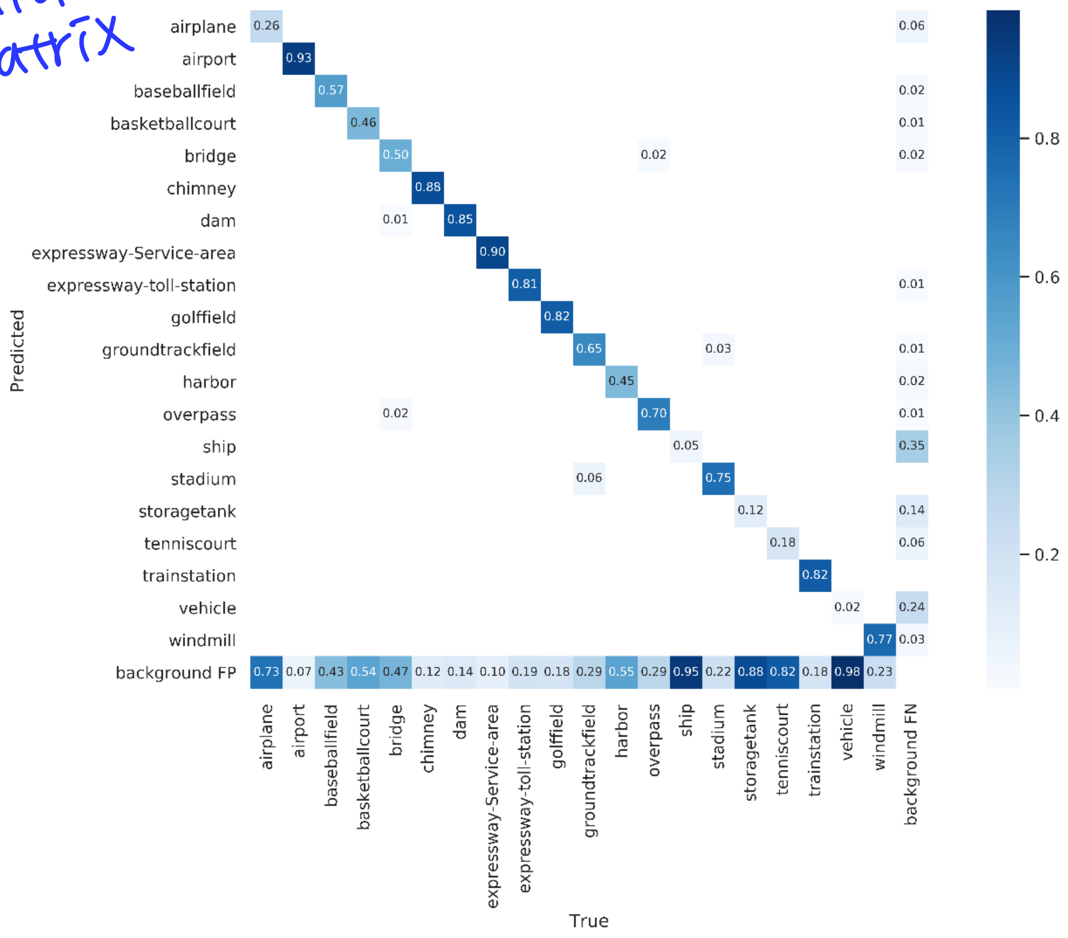
이미 사용하고 있거나 개발된 알고리즘의 간단한 확장
Supervised learning task에서 성능을 올릴수 있는 방법

Prof. Je-Won Kang
Electronic & Electrical Engineering
Ewha Womans University

Performance evaluation
in supervised learning

confusion matrix





$$\text{Accuracy} = \frac{TP+TN}{ALL}$$

Images from https://in2techs.com/confusion-matrix/?doing_wp_cron=1652020831.2095940113067626953125

Performance evaluation
in supervised learning

Prediction

Actual

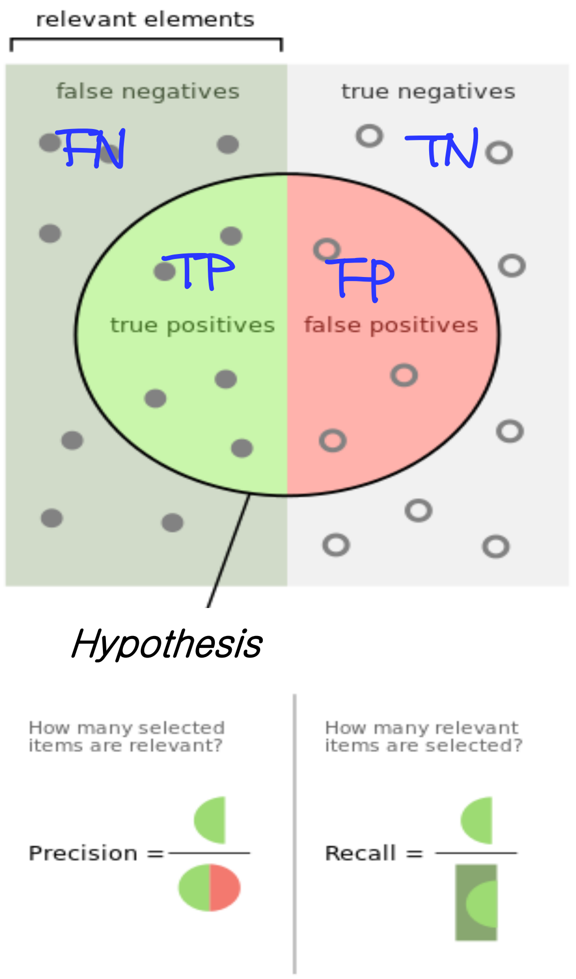
1000	정상 판정	암 판정
정상 환자	988 (TN)	2 (FP)
암 환자	1 (FN)	9 (TP)

*Confusion matrix

False positive error :
predict = positive but
actual = negative

False negative error :
predict = negative but
actual = positive

Positive samples **Negative samples**



Performance evaluation
in supervised learning

Prediction

Actual

1000	정상 판정	암 판정
정상 환자	988 (TN)	2 (FP)
암 환자	1 (FN)	9 (TP)

*Confusion matrix

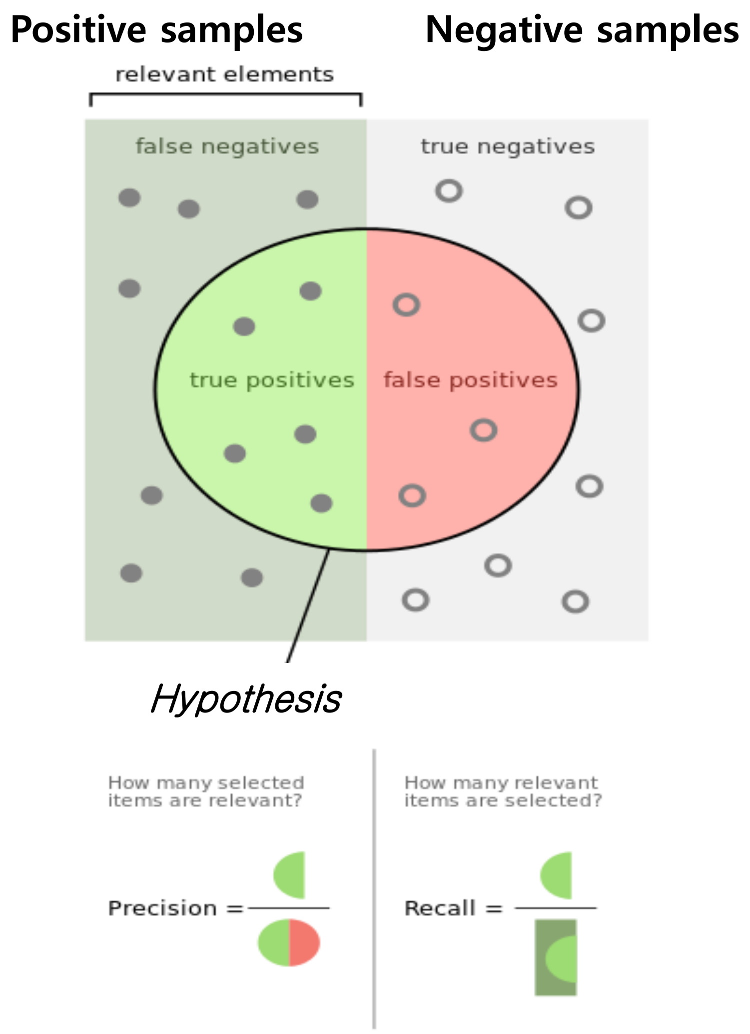
$$\text{Accuracy} = \frac{TP+TN}{ALL} = \frac{997}{1000}$$

$$\text{Precision (P)} = \frac{TP}{TP+FP} = \frac{9}{11}$$

$$\text{Recall (R)} = \frac{TP}{TP+FN} = \frac{9}{10}$$

$$\text{F1} = \frac{\mathrm{P \times R}}{\mathrm{P+R}}$$

unbalanced data의 경우 확인 필요!

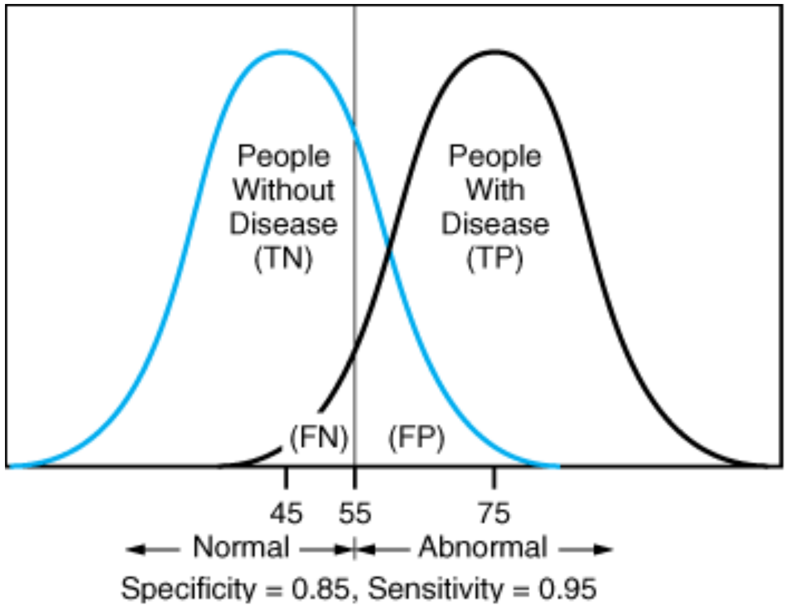


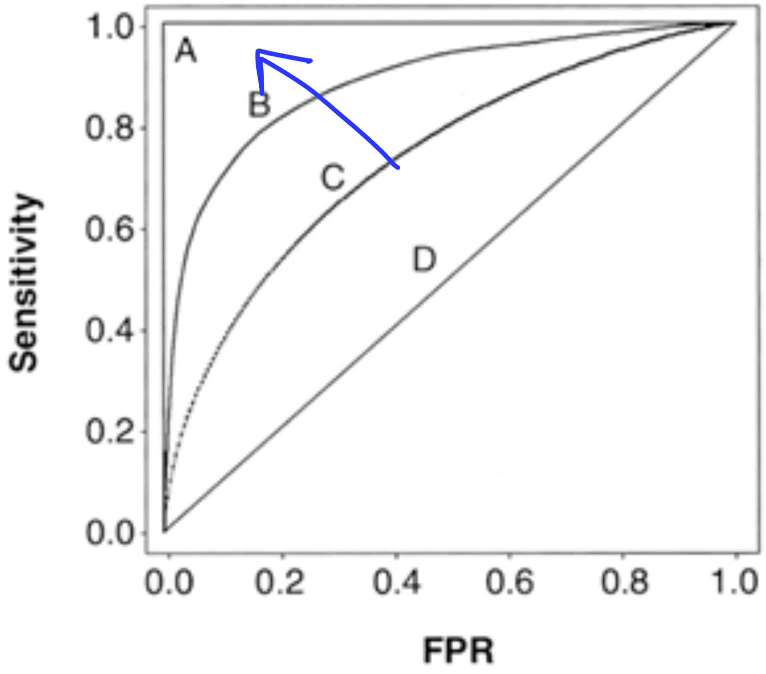
ROC Curve

- Performance comparisons between different classifiers in different true positive rates (TPR) and true negative rates (TNR).

$$TPR = R = \frac{TP}{TP + FN} \text{(recall or sensitivity)}$$

$$TNR = \frac{TN}{TN + FP} \text{(specificity)}$$





Error measure

- The error measure should be specified by the user
  - Not always given but needs to be carefully considered

Prediction

Actual

1000	정상 판정	암 판정
정상 환자	988 (TN)	2 (FP)
암 환자	$\boxed{1\ (FN)}$	9 (TP)

*Confusion matrix

recall이 높아야할것

Prediction

Actual

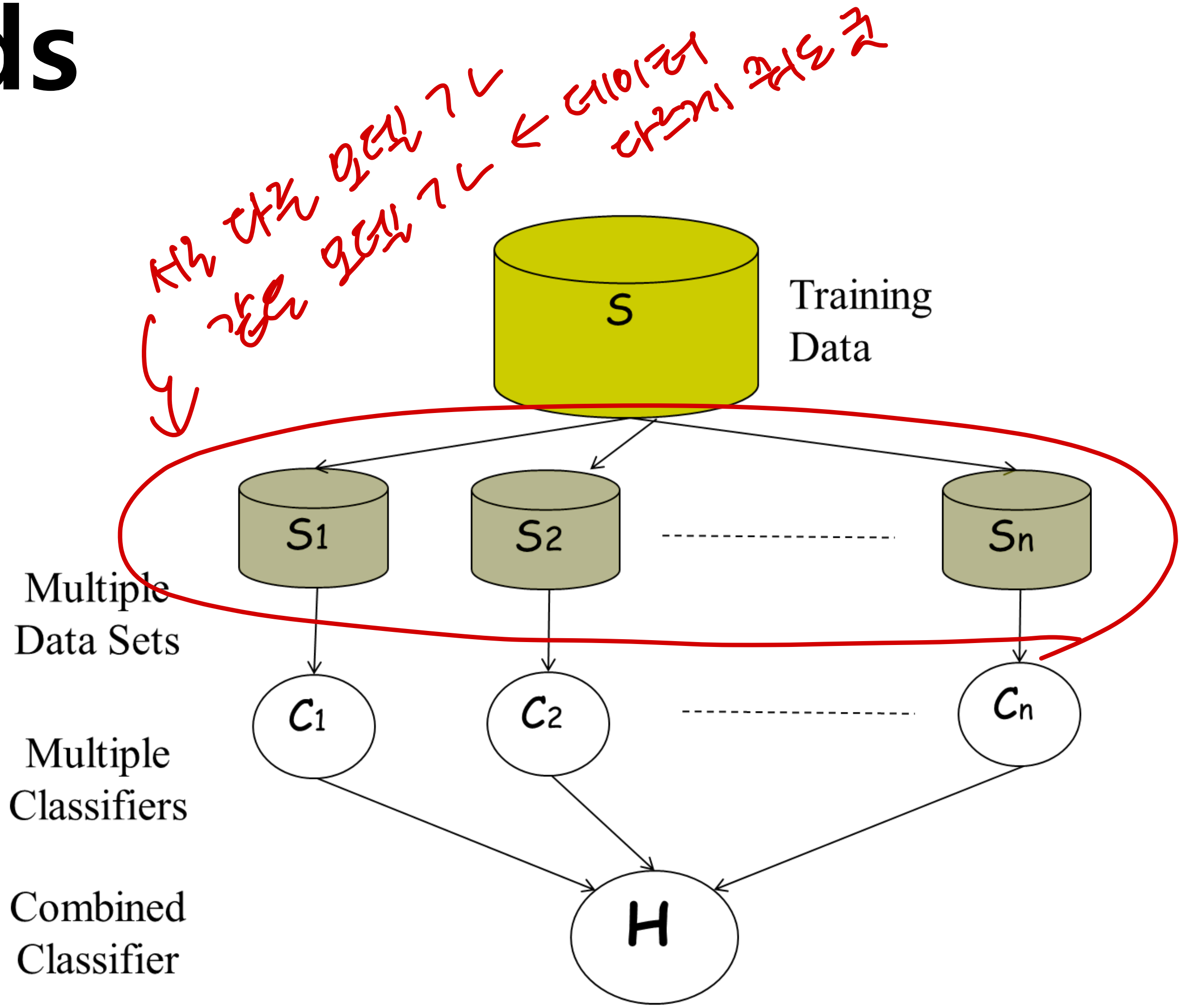
1000	지급 판정	미지급 판정
재난지원대상자	988 (TN)	$\boxed{2\ (FP)}$
재난지원비대상자	1 (FN)	9 (TP)

*Confusion matrix

precision이 높아야

Ensemble Methods

- Predict class label for unseen data by aggregating a set of predictions : different classifiers (experts) learned from the training data
- Make a decision with a voting

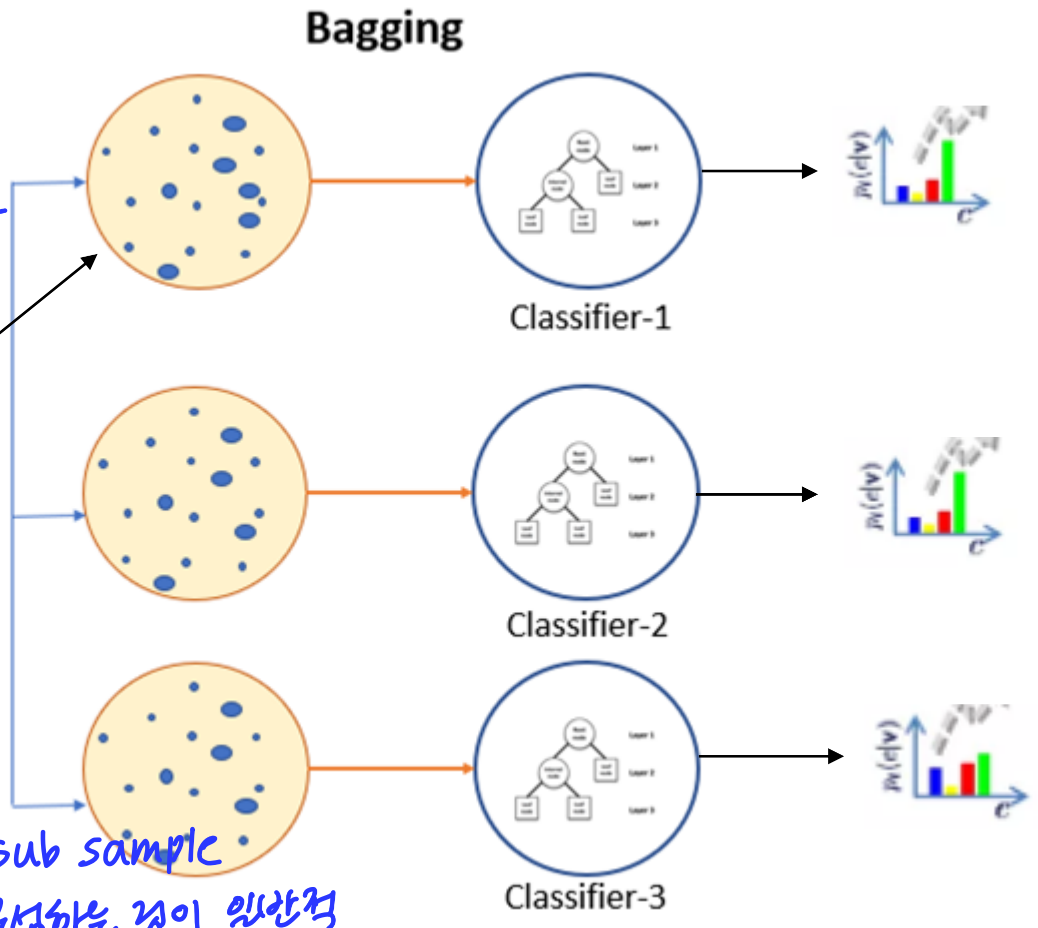


Build Ensemble Classifiers

- Basic idea: Build different experts, and let them vote.
  - Bagging and boosting
- Advantages:
  - Improve predictive performance
  - Other types of classifiers can be directly included
  - Easy to implement
  - No too much parameter tuning
- Disadvantage
  - Not a compact representation

Bagging

학습과정에서
training sample을
랜덤하게
나누어 선택해 학습

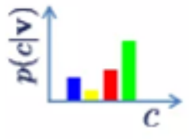


Training samples are randomly chosen

서로 다른 sub sample
구성하는 것이 일반적

The ensemble model

Forest output probability $p(c|\mathbf{v}) = \frac{1}{T}\sum_{t}^{T} p_t(c|\mathbf{v})$



or major voting

병렬학습

Images from https://www.pluralsight.com/guides/ensemble-methods:-bagging-versus-boosting

Bagging

- Bootstrapping + aggregating (for more robust performance; lower variance)
- Train several models in parallel
  - A classifier C_i is learned for each S_i in sample set S
- Bagging works because it reduces variance by voting/averaging (robust to overfitting)
  - Learning algorithm is unstable: if small changes to the training set cause large changes in the learned classifier.
  - Usually, the more classifiers the better

Bootstrapping → 다수의 sample data set을 생성하여 학습하는 방식

- Generate multiple datasets S_i in a dataset S
  - S_i has n randomly chosen samples, which may be less than the original set, with replacement
  - Repeat M times

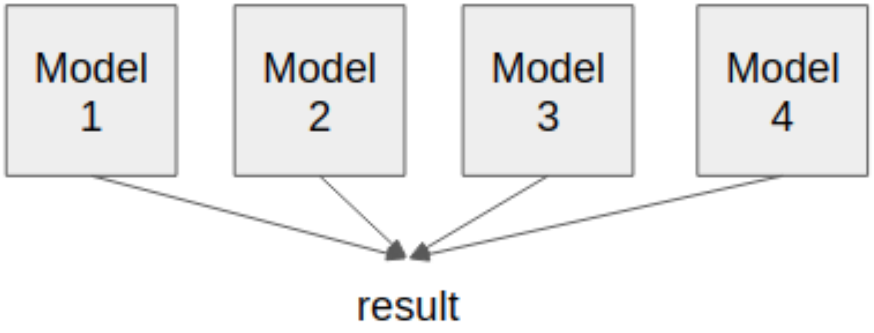
noise에 robust

→ generate M datasets, in which the size is n.

→ Train M models

Aggregating

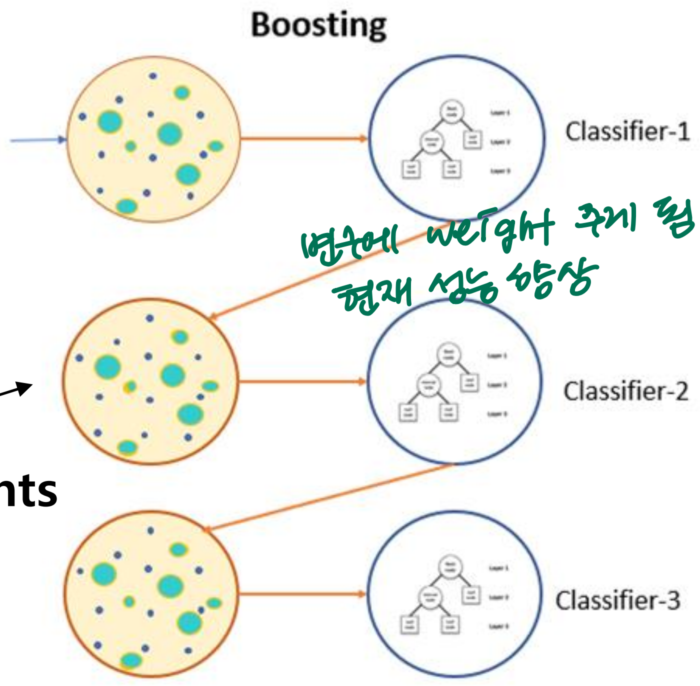
- Committee prediction



$$g(x) = \frac{1}{M}\sum_{m=1}^{M} h_m(x)$$

Boosting

- Cascading of **weak** classifiers

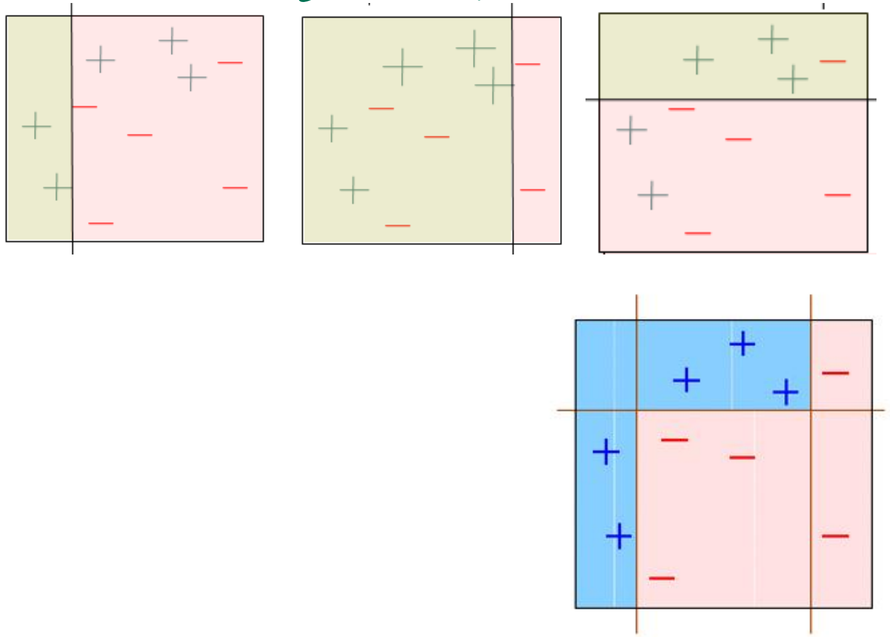


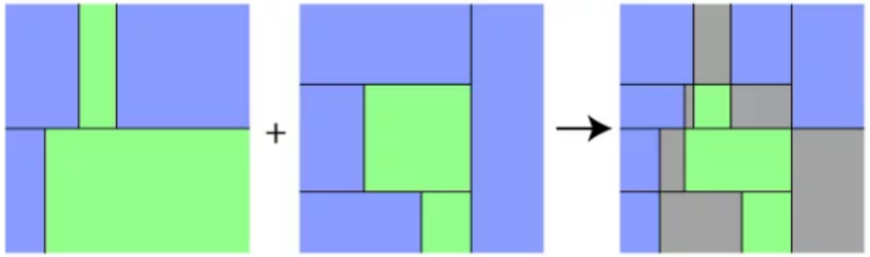
변수에 weight 주게 됨
현재 성능 향상

Chosen with weights

Sequential

week를 연속적으로 적용하게 되면





Images from https://www.pluralsight.com/guides/ensemble-methods:-bagging-versus-boosting

Boosting

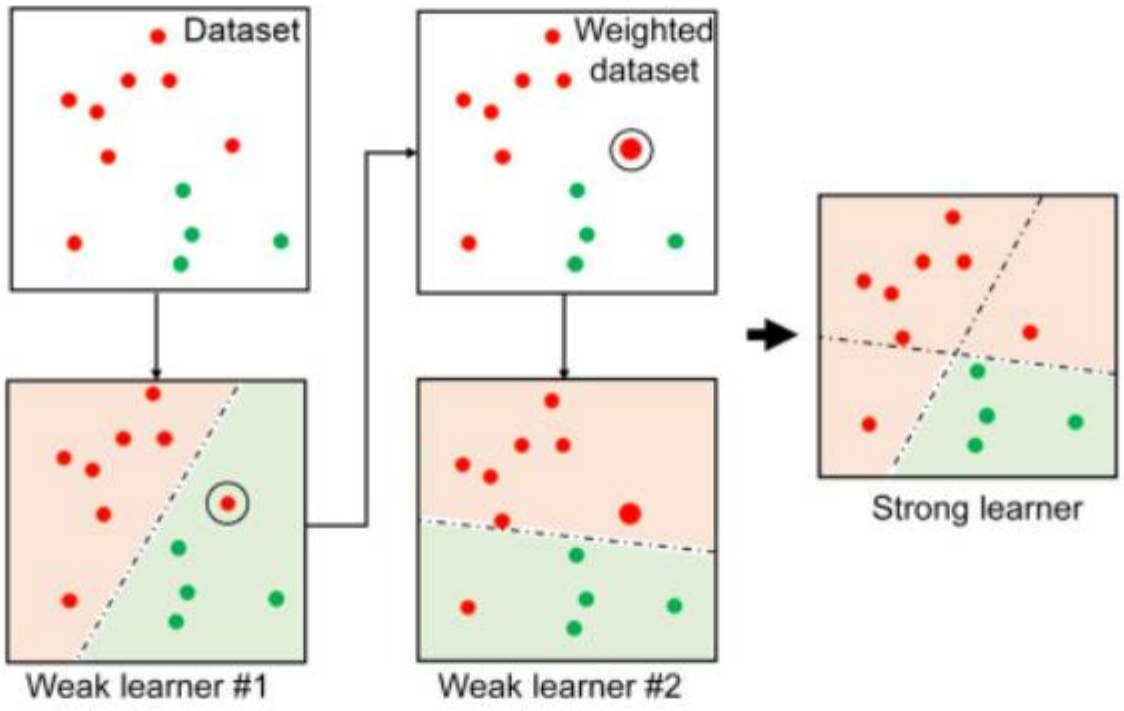
bias 높은 classifier

- Cascading of **weak** classifiers
  - Train multiple models in sequence
    - Assign a larger weight for misclassified points by one of the base classifiers, when training the next classifier in the sequence (combat to lower bias)
  - Adaboost
- Advantage
  - Simple and easy to implement
  - Flexible : can combine with any learning algorithm
  - No prior knowledge needed about weak learner
  - Versatile : can be applied on a wide variety of problems
  - Non-parametric

Adaboost

AdaBoost, short for *Adaptive Boosting, by Y. Freund and R. Shapire (1996)*

- M sequential base classifiers : $h_1, \dots, h_m, \dots, h_M$
- Trained on weighted form of the training set
- Weight depends on the performance of the previous classifier
- Combined to give the final classifier

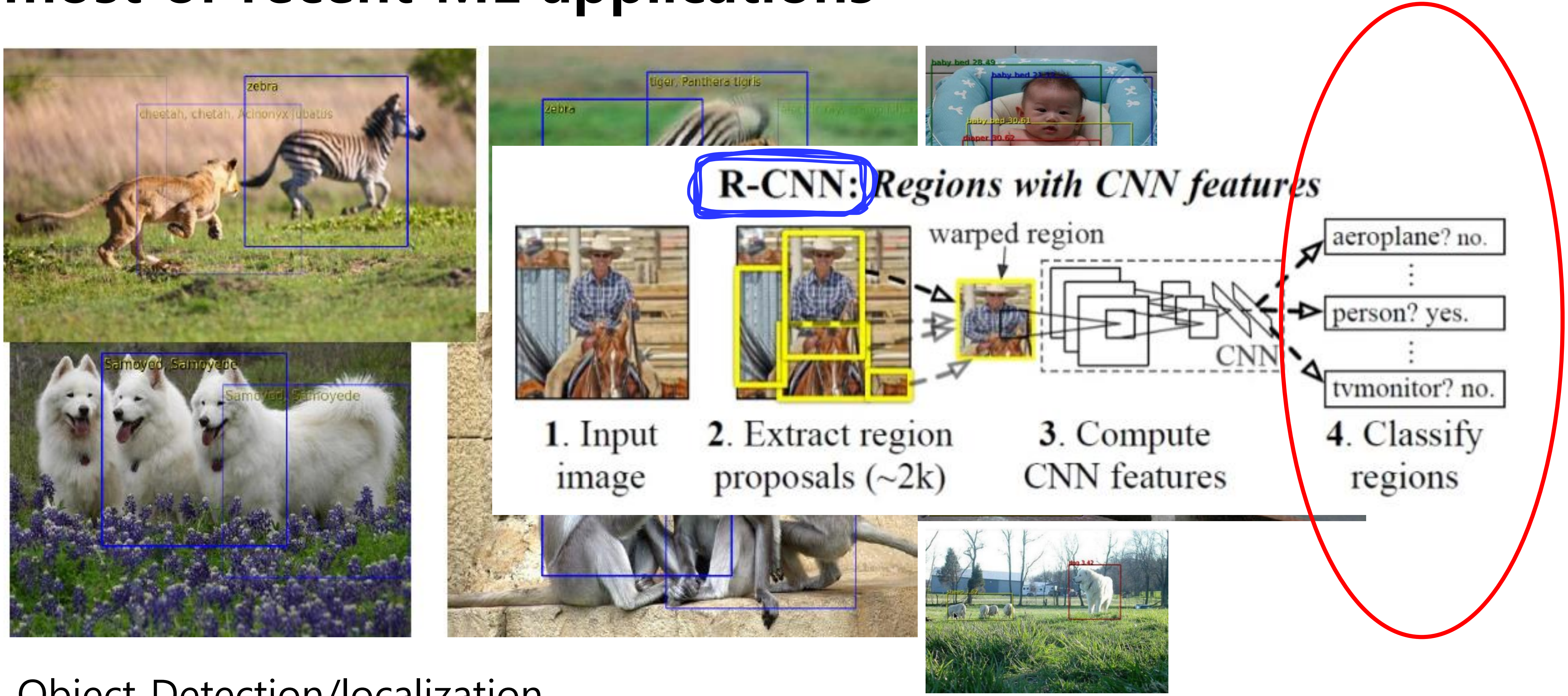


https://www.sciencedirect.com/topics/engineering/adaboost

Bagging and Boosting

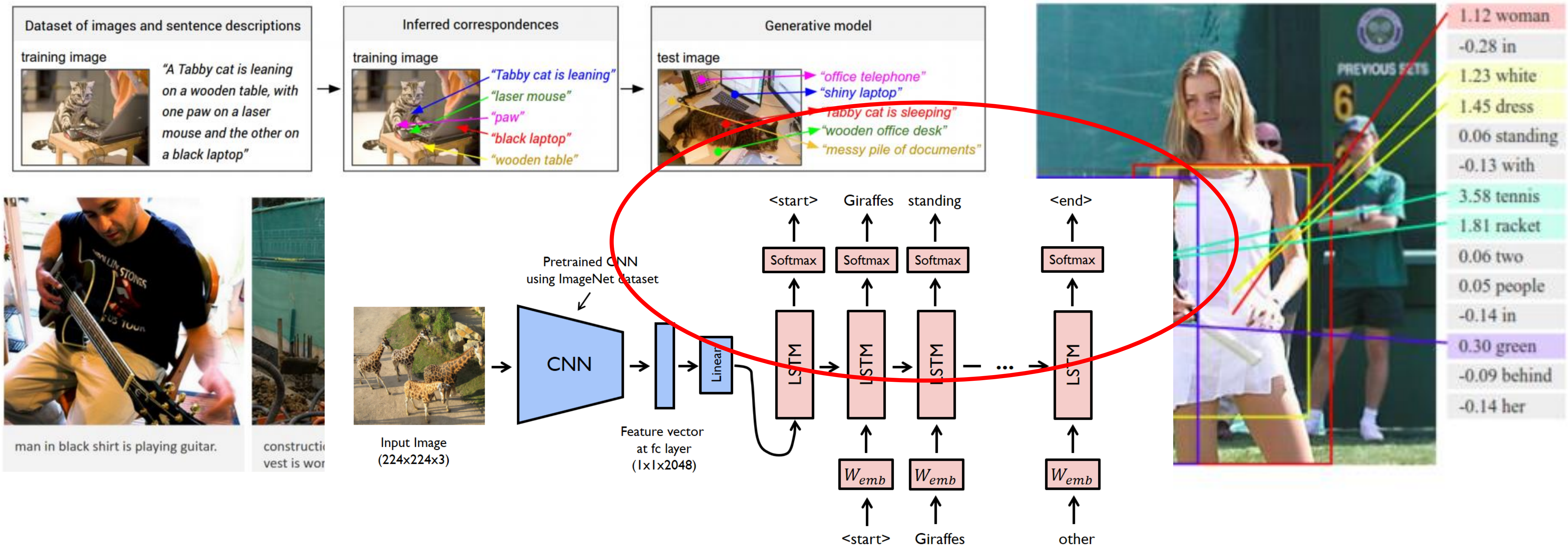
- Improving decision tree
  - By bagging -> random forest (inherently boosting)
  - By boosting -> gradient boosting machine (GBM) as generalized Adaboost
    - Very popular machine learning algorithm
    - One of leading methods for winning many Kaggle competition

Applications using SL
most of recent ML applications



Object Detection/localization

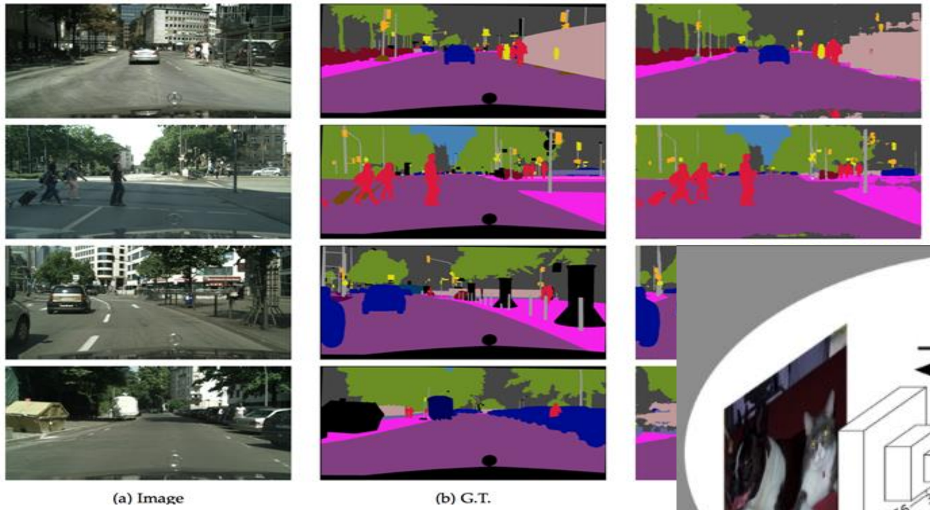
Applications using SL
most of recent ML applications



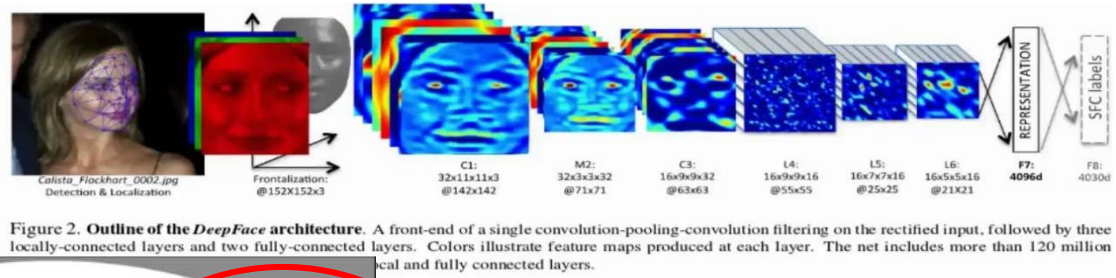
Image, language inter-disciplinary studies (image/video captioning, question and answering, etc.)

Applications using SL

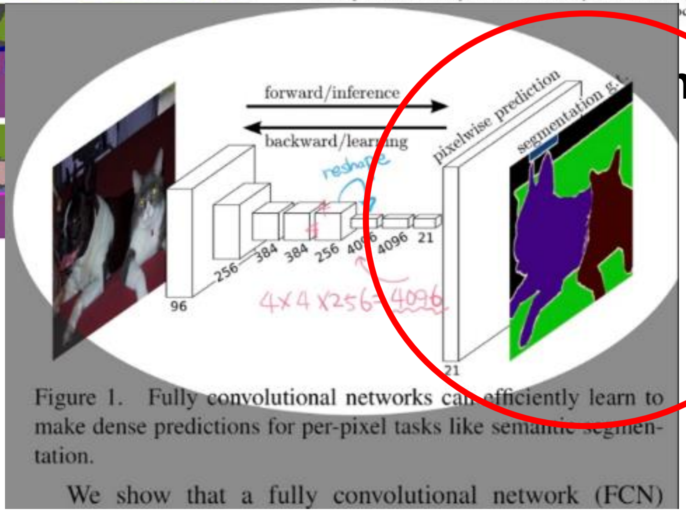
most of recent ML applications



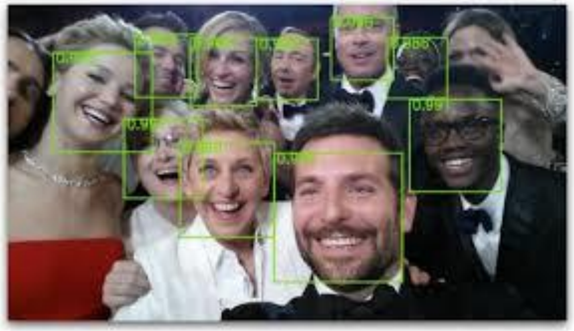
Semantic segmentation



n, pose estimation



Super resolution



Face detection

Reference

- Book: Pattern Recognition and Machine Learning (by Christopher M. Bishop)
- Book: Machine Learning: a Probabilistic Perspective (by Kevin P. Murphy)
- https://www.andrewng.org/courses/

Thank you for your attention !