

2020/10/29 토요일

- ③ Classification
- ③.1 MNIST (픽셀값 0~255) ~ 255(경) 훈련세트 선파이 (모든 고차 결승 풀과 비슷하게 훈련하기 위함)
 - ③.2 이진 분류기 훈련 $y_{train} = (y_{train} == 5) * SGD$ 은 큰 dataset 훈련하기 험하니 구조(기본 훈련 확장)
 - ③.3 선형분류
 - ③.3.1 교차검증을 사용한 정확도 측정 (Cross Validation: 훈련세트를 k개 fold로 나누어, 훈련 폴드, 테스트 폴드, 훈련 세트 확장)
 - $cross_val_score(sgd_clf, X_train, y_train_5, cv=3, scoring = "accuracy")$
 - ③.3.2 혼란 행렬 (Confusion matrix)
 - $cross_val_predict$: k개 교차검증 수행, but 평가점수가 아닌 각 test fold에서 얻은 예측을 반환.
 - y_{train_pred}
 - $confusion_matrix(y_train_5, y_train_pred) // array([[0, 5421], [0, 5421]])$ ← 원본인 분류기.

Actual \ Predicted	N	P
N (Negative Class)	TN (true negative)	FP (false positive)
P (positive class)	FN (false negative)	TP (true positive)

③.3.3 Precision (정밀도) & Recall (재현율) = $\frac{TP}{TP+FN}$ // 정확하게 감지한 양성 샘플의 비율.

ex. 감지 카메라를 통해 출도를 잡기 ($\text{민감도} (\text{sensitivity}) = \frac{TP}{TP+FN}$)

정밀도: $\frac{TP}{TP+FP}$ (인체를 인식하는 확률)

F1 Score: 정밀도와 재현율의 조화평균 = $\frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

③.3.4 정밀도/재현율 Tradeoff (정밀도를 올리면 재현율이 줄고, 반대도 마찬가지)

정밀도: 재현율 = 100% (인체를 인식하는 확률)

재현율: 100% (인체를 인식하는 확률)

③.3.5 ROC 곡선 (ROC curve) → AUC (Area Under the Curve, 훈련 데이터 히스토그램) ← (좋은 분류기: 접선으로부터 멀리 떨어져 있어야.)

FPR (False Positive Rate, 잘못 알았을 때)에 대한 TPR (True Positive Rate, 재현율, 진짜 알았을 때)의 곡선.

$= \frac{FP}{FP+TN} = \frac{(FP+TN-TN)}{FP+TN} = 1 - TN/(FP+TN) = 1 - TNR(\text{true neg rate}, 진짜 음성 비율)$

ROC 곡선은 TPR에 대한 '1-특이도(FPR)' 그래프 roc_curve(AUC 계산 함수) ← AUC (Area Under the Curve, 훈련 데이터 히스토그램)

• predict_proba() → 양성 클래스에 대한 확률을 점수로 사용. decision_function() → argmax(결과값에 대한 확률) ← OvO는 0개의 경우 반환. classes=[...].선택.

④ 다중분류 (multiclass classifier: 다중 분류기)

- OvA (one-versus-all, one-versus-the-rest, 일대다) 전략: 각 분류기 결정집합 중, 가장 높은 것을 클릭으로
- OvO (one-versus-all, 일대일) 전략: 끝나고마다 이진 분류기 훈련 (ex. 0와 1, 1과 2, ...) ← 각각 다른 class에 대한 확률

• 이미지 하나 분류하는데 45개의 분류기를 모두 통과시켜 가장 많이 양성으로 분류된 클래스를 선택.

장점 → 각 분류기 훈련에 전체 훈련 세트 중, 구별할 두 클래스에 대한 샘플만 필요함.

(SVM은 훈련 세트 크기에 인해 OvA 선택 / 대부분 이진 분류 알고리즘은 OvO 선택.)

• sklearn은 OvA, OvO 자동 선택 OneVsOneClassifier, OneVsRestClassifier로 강제 가능.

⑤ 이미지 분석 matshow()로 모히션 영상을 이미지를 표현

→ 대부분 이미지가 원래의 분류됨. ↪ 대부분 이미지가 원래의 분류됨.

diagonal 0으로 차운은 VIF(자유도)를 줄이는데 짐작.

row_sums = conf_mx.sum(axis=1, keepdims=True)

norm_conf_mx = conf_mx / row_sums

→ VIF(자유도)를 줄이는데 짐작. ↪ 1은 정확하게 분류됨.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

row_sums = conf_mx.sum(axis=1, keepdims=True)

norm_conf_mx = conf_mx / row_sums

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향에 대한 통찰을 얻을 수 있다.

→ VIF(자유도)를 분류기의 성능 향상 방향

2020 6/10 정하연

관련 아이디어: 두 클래스를 구분하는 경계(경계선) 사이의 이동을 가능한 크게 하는 것이 목표.
SVM (Support Vector Machine): 보조한 분류기(분류기), 작거나 중간 크기의 dataset.
5.1 선형 SVM 분류
 → 2차원 마진 분류(클래스 사이 넓은 품 찾기).
 · 서포트 벡터: 2로 경계에 위치한 샘플(예: 경계 경계). * 마진 벽(최대한 멀리 떨어져 있음).
5.2.1 소프트 마진 분류 → 하드 마진 분류(2로 샘플이 분류됨). * 선형 세트의 Scale이 맞지 않으면 크기가 다른 특성을 사용하는 방향.
 도수로 평가해보면 미분이 더 좋다. → 데이터에 따라 구분되야. (이상화에 영향).
 (정답률은 datapoint) → 데이터의 정답률 찾기.

sklearn: C파라미터 → 두 가지 마진 유형(SVM의 고려 대상일 때 (\downarrow 오류제 = 편향)).
 (고려할 때마다 고정 정도). test samples과 경계를 사이의 거리를 측정할 수 있으므로 선형으로 접근 가능.
 *[LinearSVC(C=1, loss="hinge")]

시간	predict_proba()	특징	최종화 안내
dense API	LinearSVC	• 경계경계 x	• 확률수행
scikit-learn API	SVC	• margin Δ(probability=True)	• 차운, 정규 dataset • 아래에 표현으로 loss func
기본함수	SGDClassifier	o	• loss, 학습률 • SGD는 loss func

부록 예제의 종결.

5.2 비선형 SVM 분류: Polynomial feature와 같은 특성을 추가. → \uparrow \uparrow \uparrow
poly_svm_clf = Pipeline([PolynomialFeatures(degree=3), StandardScaler(), LinearSVC(c=10, loss="hinge")])

5.2.1 다항식 커널 (낮은 차수: 복잡한 데이터셋 표현 X, 높은 차수: 느림)

- 커널 트리(kernel tree): 실제로 특성을 추가하지 않고서도 다양한 특성을 양이 추가한 것 같은 결과.
- poly_kernel_svm_clf = Pipeline([StandardScaler(), SVC(kernel="poly", degree=3, C=100, loss="hinge")])**
- 모델이 과대적합이면 차수 ↓, 과소적합이면 차수 ↑ (과대 \rightarrow 차수 10, 차수 1, C=5)) 차수에 의해 영향을 받을 것인가.

5.3 SVM 학습

- 선형 비선형 SVM 분류와 유효한 반대로.
- 제한된 마진으로 인해서 고도 압축 가능한 많은 샘플이 들어가도록. (도수: 단조로운, 크면 느림).
 (도수 높은 샘플)
- SVC** | C 조절로 $\begin{cases} \text{C} \uparrow \text{유체능}, \\ \text{C} \uparrow \text{유체구} \end{cases}$ → 마진 편평. ← 큐브 ← ↑
SVR | 마진으로 인해 penalty 조절 $\begin{cases} \text{C} \uparrow \text{유체능}, \\ \text{C} \uparrow \text{유체구} \end{cases}$ → C 조절로 ↑
 : 마진으로는 충분한 샘플 추가되어도 모델의 예측에는 영향 X)
- (선형) **svm_reg = LinearSVR(epsilon=1.5)** → 훈련세트의 편평 시간.
- (비선형) **svm_poly_reg = SVR(kernel="poly", degree=2, C=100, epsilon=0.1)** → 훈련세트↑. 편평성이 비선형적으로 증가.

1 Ensemble Learning, Random Forest

경계트리의 악상을 (일반화로 배깅 or 페이스팅 적용).
 · 일반화 예측기로부터 예측은 수집하면 가장 좋은 모델보다 더 좋은 예측을 얻을 수 있음.

· 암상을 학습: 암상을 학습 알고리즘 (백정, 페이스팅)

7.1 Voting classifiers (투표 기반 분류기) (전형화보다 2배 더 높은 성능을 내는 분류기). * 투표 양수를 더 나은 투표수(5개로 5개로)

· hard voting (강한 투표): 각 분류기의 예측을 모아서 대체로 투표로 클래스로 예측. * 각 분류기가 '약한 학습기' (weak learner) 일지라도 충분하게 많고 다양하다면 암상을 높은 정확도로 예측.

· soft voting (경계트리): 개별 분류기의 예측 확률을 평균내어, 확률이 가장 높은 클래스로 예측. $\rightarrow \text{SVC probability} = \frac{\text{True}}{\text{True} + \text{False}}$

7.2 Bagging + Boosting → 선형으로는 + 소량 편평.

· 다양한 분류기 만들기 방법 ②: 훈련세트의 Subsample를 무작위로 구성하여 분류기를 각기 다르게 학습시킴 → Boosting: 편평 X.

· 편평한 분산↑ (vs) 1개 예측기 (편평 배우, 분산은 1개가)

7.3 랜덤 포치 + 랜덤 subspace

Bagging Classifier는 특성 샘플의 차원. → 각 예측기는 무작위로 선택된 입력 특성의 일부로 훈련.

· 랜덤 포치 방식: 훈련 특성과 샘플 모두 샘플링. (patches method)

· 랜덤 subspace 방식: 특성 샘플(bootstrap_features=True, max_features=1.0) → ex. 불꽃 dataset 특성 수에 대해 고정하고.

7.4 Random Forest

RandomForestClassifier(n_estimators=500, max_leaf_nodes=16, n_jobs=-1) → 모든 CPU를 사용.

Extreme Randomized Tree (편평한 분산) n_estimators=500, max_samples=1.0, bootstrap=True, n_jobs=-1

<RandomForest>: 랜덤 훈련, 전신의 일정한 (어떤 알짜기 기준으로 노드를 분할하기). → 예측 시간에 차이는 없음.

ExtraTreesClassifier: 각각 특성, 전신 학습법(구체화하기 직전) / 랜덤 포레스트가 고대적합이면 엑스트라 트리는 X / 전신학습법 일정에서 출현 비율.

7.4.2 특성 중요도 (feature importances) = 가중치 평균 (노드의 가중치 = 평균된 훈련 샘플 수).

: 노드에서의 복수도 감소가 그린 해상 노드의 특성이 충분하다고 간주되며, 이 줄어드는 해당 노드에 속한 훈련 샘플 수 ↓

→ (2 노드 (노드의 가중치) × (감소하는 복수도)) / (전체 가중치의 합) ≈ 0.1 사이의 값(이 가중치를 줄여)

→ 랜덤 포레스트는 특성을 선택해야 할 때 어떤 특성이 충분히 빠르게 파악할 수 있어 매우 편리.

(각 DT의 특성 중요도를 쉽게 평균화. 편평 가능).

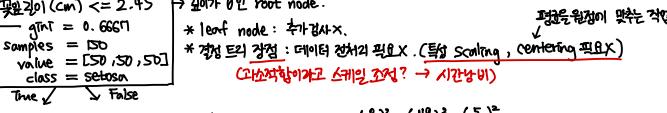
11 연습문제

· 배깅 페이스팅 방향을 & 전신 포레스트 → 각 예측기가 독립적이며 여러 대의 서버에 분산시켜 속도를 높일 수 있음.

· oob 평가의 장점? 추가 검증세트가 필요없어서, 훈련에 더 많은 샘플 사용 가능 → 성능의 상승 보장 가능.

6 DT (Decision Tree): 결정트리 → 분류, 회귀 & 다중분류. / Random Forest의 기본 구성 요소.

6.1 결정트리 학습 및 시각화

tree_clf = DecisionTreeClassifier(max_depth=2) // 결정트리 2. 
 * export_graphviz(tree_clf, out_file="tree.dot", ...)
 * 결정트리의 경계: 데이터 분리에 필요한 특성 (특성 Scaling, centering 필요).
 고려사항(부록) → 스키마 조정? → 시간상비)

꽃잎 2cm (cm) <= 2.45 → 꽃잎이 0인 root node.

꽃잎 2cm (cm) > 2.45 → 꽃잎이 1인 leaf node.

* leaf mode: 추정값 X.

* 결정트리의 경계: 데이터 분리에 필요한 특성 (특성 Scaling, centering 필요).

꽃잎을 원점에 맞춘 적당한 경계.

꽃잎을 원점에 맞춘 적당한 경계.