인공지능 과제1

\* 문제 아래에 본인의 답을 서술하고, “코드제출”이라고 적힌 문제(3, 4, 5, 7, 8, 10번)에 대해서는 코드를 *Assignment.zip* 파일로 압축하여 따로 제출하시오.   
(단, 코드 파일명은 *Assignment-문제번호.py* 로 할 것)

01. 순서형 특징과 이름형 특징의 사례를 두 가지씩 제시하시오.

**순서형 특징의 사례: 학점, 수능등급**

**이름형 특징: 혈액형, 지역**

02. ◆은 정상품이고 ◇은 불량품이다. 생산 라인에 설치된 자동 검사 시스템이 12개 생산품에 대해 다음과 같이 P 또는 N으로 예측했다. P는 불량, N은 정상이다. 이를 통해 (1)~(3)의 결과를 적으시오.

◆N ◆N ◆P ◇P ◆N ◆N ◇P ◆N ◆N ◆N ◇N ◆N

(1) 정확률은 얼마인가? **10/12=5/6=83.3%**

(2) 정밀도와 재현율은 얼마인가? **정밀도:2/2=3=66.7% 재현율:2/3=66.7%:**

(3) 특이도와 민감도는 얼마인가? **특이도:8/9=88.9% 민감도: 2/3=66.7%**

03. Assignment-3.py 파일을 참조하여 SVM뿐만 아니라 결정 트리 모델까지 적용하도록 확장하시오. 이때 훈련 집합을 테스트 집합으로 간주해 SVM과 결정 트리의 정확률을 측정하여 적으시오. -코드제출

[hint] 결정 트리는 https://scikit-learn.org/stable/modules/tree.html을 참조한다.

SVM 정확률= 98.66666666666667 %

DTC 정확률= 100.0 %

04. Assignment-4.py 참조하여 SVM, 결정 트리, 랜덤 포리스트의 세 가지 기계 학습 모델 중에 어떤 것을 쓸지 결정하는 모델 선택을 해야 한다고 가정하고 5-겹 교차 검증을 사용하여 모델 선택을 수행한다. 모델을 선택하고 선택된 모델을 가지고 테스트 집합에 대한 성능을 측정하여 출력하는 프로그램을 작성하시오. 이때 sklearn이 제공하는 digit 데이터셋을 사용하시오**.** (훈련 집합에 대한 세 모델의 평균 정확률을 적고, 그 중 가장 성능이 좋은 모델을 선택하여 테스트 집합에 대한 정확률을 적으시오.) -코드제출

[hint] 결정 트리는 https://scikit-learn.org/stable/modules/tree.html을 참조한다. 랜덤 포리스트는 https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html을 참조한다.

SVM 정확률(평균)=98.051

DTC정확률(평균)=82.283 %

RFC 정확률(평균)=82.005 %

선택된 모델 SVM

정확률= 99.72183588317107 %

05. 4-5.py를 MNIST에 적용하시오. 최적의 은닉층의 노드 개수와, 테스트 집합에 대한 정확률을 적으시오. (Assignment-5.py 참조) -코드제출

코드 실행시간이 너무 오래 걸리므로,

* max\_iter=300 🡪 20
* cv=10 🡪 5
* prange=range(50,1001,50) 🡪 range(50,1001,200)

으로 수정해서 돌릴 것.

최적의 은닉층의 노드 개수는 850 개입니다.

테스트 집합에 대한 정확률은 96.21428571428572 %입니다.

06. MLPClassifier 함수에서 shuffle 매개변수의 역할을 설명하시오.

**새로운 세대를 시작할 때 마다 훈련 집합에 있는 샘플 순서를 뒤집는 역할을 한다.**

07. 4-4.py는 은닉층이 하나인 다층 퍼셉트론을 사용한다. 은닉층을 2, 3, 4, 5개로 점점 늘려 깊게 만들면서 측정한 정확률을 모두 적고, 이렇게 깊어지면 성능 향상이 있는지 서술하시오. 실험은 MNIST 데이터를 사용하시오. (Assignment-7.py 참조) -코드제출

(은닉층 1개) 테스트 집합에 대한 정확률은 97.78999999999999 %입니다.

(은닉층 2개) 테스트 집합에 대한 정확률은 97.91 %입니다.

(은닉층 3개) 테스트 집합에 대한 정확률은 97.98 %입니다.

(은닉층 4개) 테스트 집합에 대한 정확률은 98.08 %입니다.

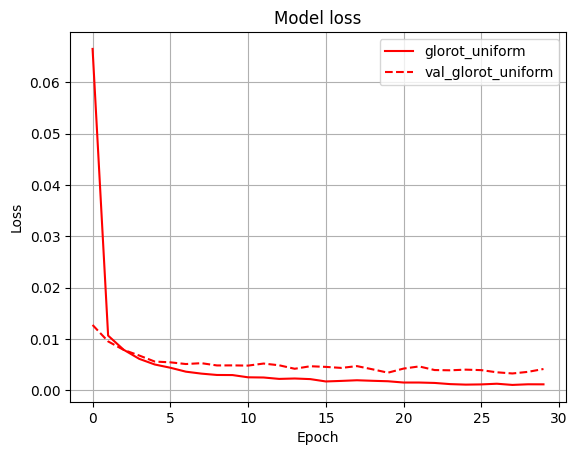
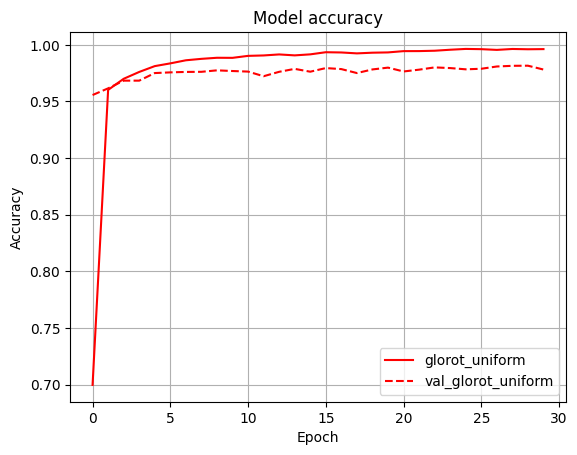
(은닉층 5개) 테스트 집합에 대한 정확률은 98.07000000000001 %입니다.

은닉층이 깊어질수록 대체로 성능이 향상된다는 사실을 알 수 있다.

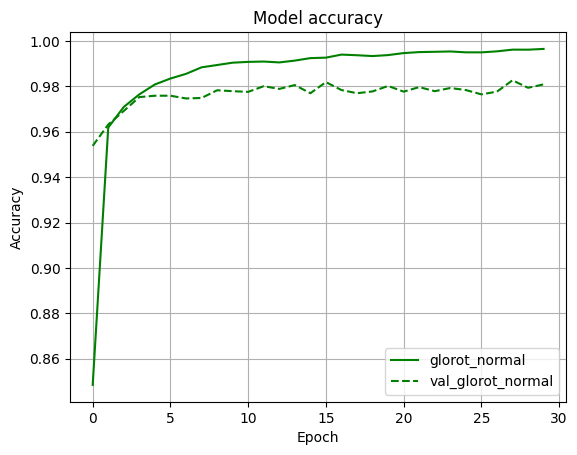
08. 5-9.py의 27~31행에 있는 가중치를 초기화하는 매개변수를 glorot\_uniform과 glorot\_normal, random\_normal로 바꾸면서 성능을 비교하시오. 세 결과의 정확률을 적고, 각 결과에 대한 그래프를 붙여넣으시오. -코드제출

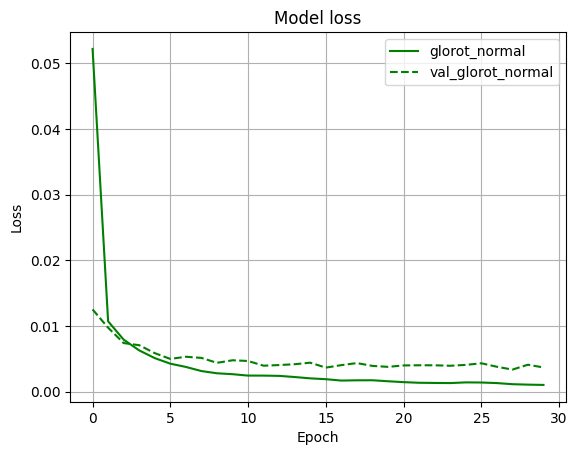
glorot\_uniform

정확률은 97.82000184059143 %

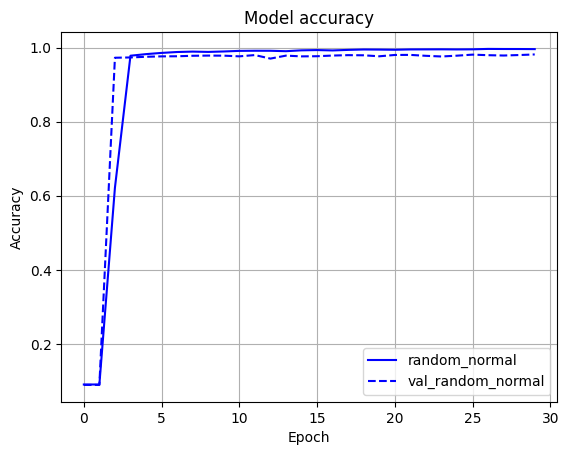


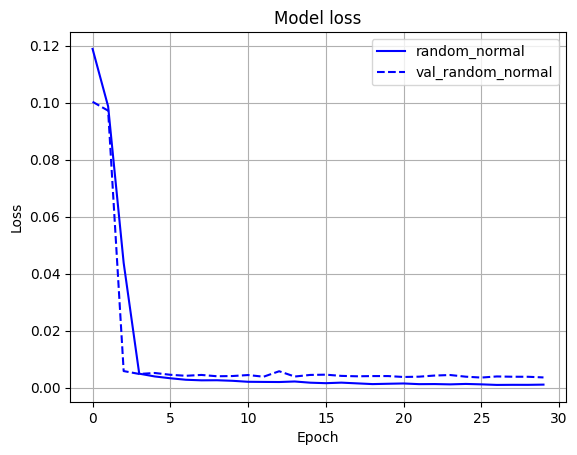
glorot\_normal

정확률은 98.089998960495 %



random\_normal

정확률은 98.15999865531921 %



09. 샘플의 부류 레이블이 y=(0,0,1,0)인데 모델이 o=(0.01,0.09,0.8,0.1)을 예측했을 때의 교차 엔트로피를 계산하시오. 또 다른 모델이 o=(0.97,0.01,0.01,0.01)을 예측했을 때의 교차 엔트로피를 계산하시오. 그리고 두 경우를 비교해 교차 엔트로피의 합리성을 설명하시오.

[hint] [예제 5-1]을 참조한다.

**y=(0,0,1,0)일 때 모델이 o=(0.01,0.09,0.8,0.1)인 경우는 부류 3에 해당하는 곳이 0.8로서 가장 크므로 신경망이 샘플을 맞힌 경우이다. 여기서 교차 엔트로피 값을 구해보면**

**의 값을 가진다. 모델이 o=(0.97,0.01,0.01,0.01)인 경우는 부류0에 해당하는 곳이 0.97로 가장 크므로 신경망이 틀린 경우이다. 이 경우 교차 엔트로피 값을 구해보면**

**의 값을 가진다. 이를 통해 교차 엔트로피는 틀린 경우에 더욱 큰 값을 발생시키기 때문에 손실 함수로 사용하기 적합하다.**

10. 5-12.py는 fashion MNIST 데이터셋에 대해 90% 가량의 정확률을 얻었다. 5장에서 배운 기법을 동원해 다층 퍼셉트론의 성능을 최대한 끌어 올리시오. 이때 성능 측정은 교차 검증을 이용하시오. 이렇게 끌어올린 자신의 정확률과 하이퍼 매개변수 값을 제시하시오. -코드제출

사용된 하이퍼 매개변수

학습률: 0.001000 옵티마이저: Adam, 배치사이즈: 512, epoch: 80

최적의 정확률: 98.11500072479248 %