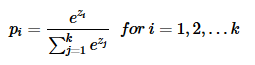
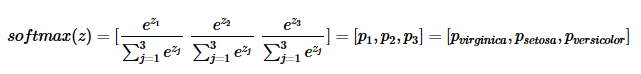
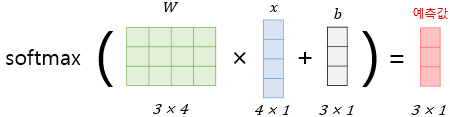
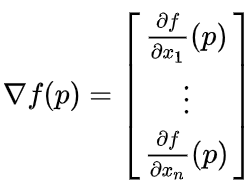
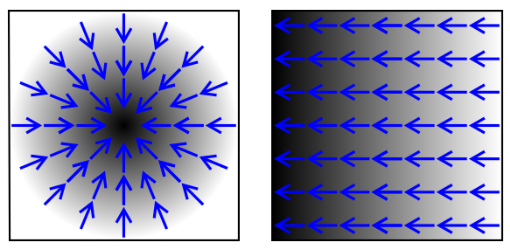
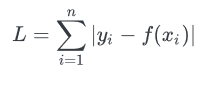
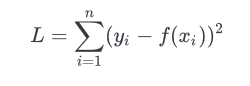
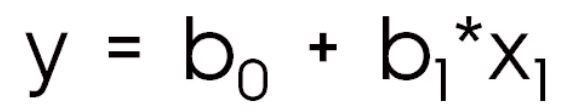
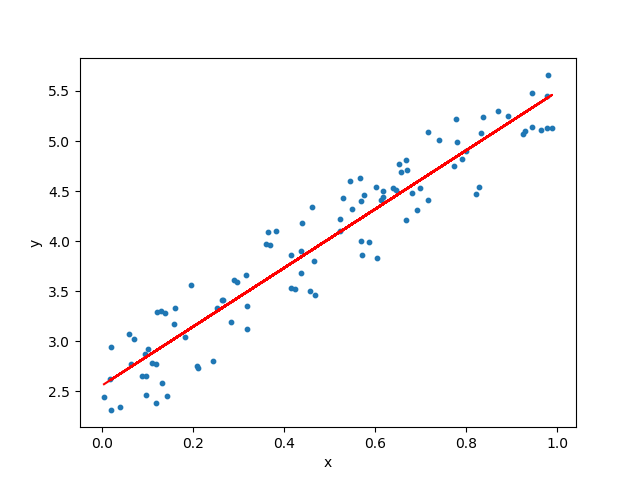
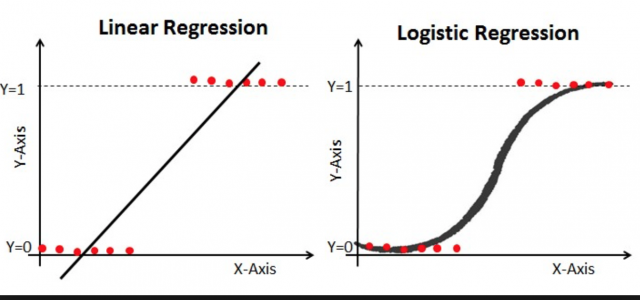
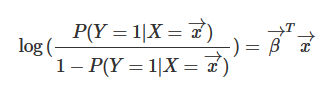
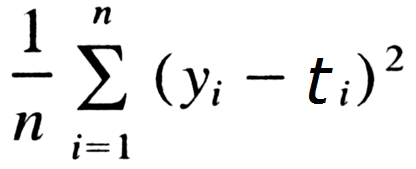
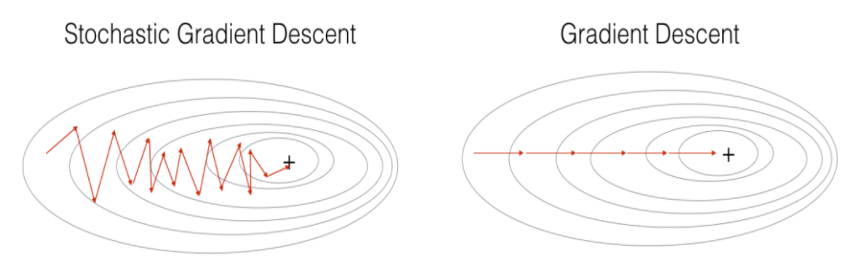
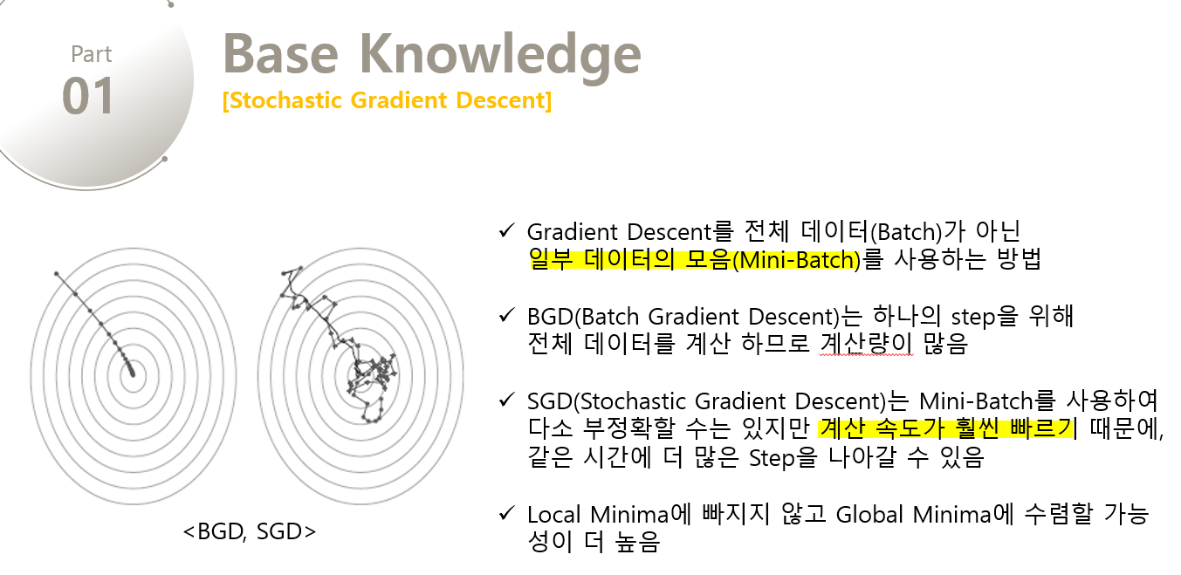
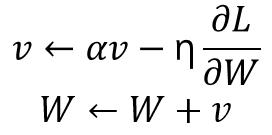
1. Softmax
   1. 각 클래스의 확률을 구하는 함수
   2. 확률의 합은 정확히 1.0
   3. 
   4. k 차원의 벡터에서 i번째 원소를 z\_i, i번째가 정답일 확률을 p\_i
   5. 
   6. 입력백터 x, 가중치 행렬 W, 편향 b
   7. 
2. Gradient
   1. 미분 가능한 f가 여러 번수에 의해서 정희된다고 했을 때 포인트 p에서 f에 대한 편미분
   2. 
   3. 
   4. 검정색이 실제 함수의 값이 높은 것을 의미하면 Gradient는 화살표로 나타낸다
3. Heuristic
   1. 경험에 기반하여 문제를 해결하거나 학습하거나 발견해내는 방법
   2. 휴리스틱 알고리즘
      1. 자원제한적인 상황으로 인해 최적의 해를 구하는 것이 현실적으로 불가능한 경우, 알고리즘의 목적 중 하나 또는 둘 모두를 포기하면서 상황내에서의 최선의 해답을 찾는 알고리즘
      2. 일단 답을 내지만 정확만 답을 보장할 수 없다.
4. L1 loss
   1. 실제 값 y\_i 와 예측값 f(x\_i) 사이의 차이값에 절대값을 취하고 모두 합한 것
   2. Least Absolute Deviations, LAD라고도 함.
   3. 
   4. 0인 지점에서 미분이 불가능하다.
5. L2 loss
   1. 실제값과 예측값 사이의 오차를 제곱한 값들을 모두 합한 것
   2. 
   3. Least Square error, LSE라고도 함.
   4. 미분이 가능하다는 장점이 있다.
   5. 비교
      1. 이상치가 등장했을 때 L2가 더 민감하게 반응한다. Outliers 가 효과적으로 적당히 무시되길 원한다면 L1 loss를 골라야 한다.
6. Linear regression
   1. 독립변수와 종속변수에 선형적인 관계가 있다고 가정
   2. 간단한 선형 방정식
   3. 
   4. 
   5. **Hypothesis**
   6. *hθ*(*x*)=*θ*0+*θ*1*x*
   7. 정확도를 구하기 위해서 cost function을 이용한다. MSE를 이용한다.
7. Logistic regression
   1. 
   2. 확률의 성질을 만족하도록 regression function을 보완하는 방법론
   3. Logistic function
   4. 
   5. 로지스틱 회귀 모델의 식
   6. 
8. MSE
   1. 평균제곱오차라고 부른다.
   2. 
   3. 제곱 손실을 예시의 개수로 나뉘어 계산한다.
   4. Cost function으로 많이 쓰인다.
   5. 값이 작을수록 더 좋은 성능을 보인다.
9. SGD
   1. 
   2. 추출한 데이터 한 개에서 error gradient를 계산하고 gradient descent를 적용하는 방법
   3. Local optimal에 빠질 위험이 적다
   4. Step에 걸리는 시간이 짧기 때문에 수렴속도가 상대적으로 빠르다
   5. Global optimal을 찾지 못할 수 있다
   6. 데이터를 한 개씩 처리하기 때문에 gpu 성능을 전부 활용할 수 없다.
   7. Gradient Descent에서 Loss function을 계산하게 된다.
   8. 전체 Train-set을 사용하는 것을 Batch Gradient Descent라 하는데 전체 데이터에 대한 Loss Function을 구해야 해서 계산양이 많이 필요하다. 전체 배치 대신에 Mini batch를 사용해서 더 빠르게 step을 갈 수 있다.
   9. 
10. Momentum
    1. Gradient descent 기반의 optimization algorithm
    2. 
       1. L : Loss function value
       2. W : weights
       3. n : learning rate
       4. a : hyper parameter
       5. v : 일종의 가속도
    3. 공 굴리듯이 가중치가 이동한다.
    4. 