|  |  |  |
| --- | --- | --- |
| 12주차 일석이조 조별보고서 | | |
| 작성일 : 2023년 11월 21일 | | 작성자 : 이준용 |
| 조 모임 일시 : 11월 21일 | | 모임 장소 : 구글meet |
| 참석자 : 이준용, 유정훈, 김동규, 이학빈, 탁성재 | | 조원 : 이준용, 유정훈, 김동규, 이학빈, 탁성재 |
| 구분 | 내용 | |
| 학습 범위와 내용 | 12주차 온라인 강의 내용  교재 5장 내용 | |
| 논의 내용  (조별 모임 전에 조장이 지시) | 기계학습 5장 딥러닝 최적화  5.3 규제의 필요성과 원리  5.4 규제 기법  5.5 하이퍼 매개변수 최적화  5.6 2차 미분을 이용한 최적화 | |
| 질문 내용  (모임 전 공지된 개별 학습 범위에서 이해된 것과 못한 것) | **Q1**  Norm 과 Metric의 차이에 대하여 알고 싶습니다.  **A1**  Metric은 학습을 통해 목표를 얼마나 잘 달성했는지를 나타내는 척도이다. 여기서는 한 가지 손실 함수만을 다룬다. 훈련을 계속하다 보면 손실 값이 줄어들면서 척도 값도 줄어들지만, 정말 수렴할 때 즈음에 이르러서는 손실 값이 줄어도 척도 값이 줄지 않기도 한다. 이는 둘의 계산식이 다르기에 일어나는 현상이다.  그래서, 훈련 막바지에 검증 데이터에서 손실 값이 줄어드는 것을 모니터링 하지 않고 척도 값이 줄어드는 것을 모니터링 한다.  Norm은 벡터의 길이 혹은 크기를 측정하는 방법이다. Norm이 측정한 벡터의 크기는 원점에서 좌표까지의 거리 혹은 크기라고 한다.    으로 나타낼 수 있고, p는 Norm의 차수를 의미한다.  P가 1이면 L1 Norm이고 p는 2이면 L2 Norm이다.  N은 벡터의 요소 수이다. Norm은 각 요소별로 요소 절대 값을 p번 곱한 값의 합을 p재곱근한 값이다.  **Q2**  뉴턴 방법에 대해서 궁금하다.  **A2**  2차미분의 최적화 방법 중 요소 및 가중치를 m개를 가진다고 정의된 벡터는 뉴턴 방법을 이용해 최적화를 시킬 수 있습니다.  예를 들어, f(x)=x^7, x=a의 값을 넣었다고 가정할 때, 해당 |a|가 크면 클수록 점점 그래프가 넓어질 것 입니다. 또한 기울기는 완만할 것입니다. 하지만 역으로 |a|가 작다면 그래프는 좁고 가파를 것입니다.  뉴턴 방법은 해당 a를 계속 이동시키며, 접선과 x축이 만나는 지점으로 계속 이동합니다.    2차 미분 값은 해당 f(x)=(x-xk)^n을 2번 미분한 값으로 표현될 수 있습니다.  이때 f’(x), f’’(x)은 각자 미분값을 표현하는 헤시안 행렬 나타납니다. 하지만 이때 m개의 요소를 가지면 m\*m 크기의 행렬을 가지게 되므로, 켤레 그래디언트, 유사 뉴턴 방법등을 사용합니다.  켤레 그래디언트는 현재 계산하는 위치가 아닌, 바로 직전 위치에서의 그래디언트를 활용합니다.  유사 뉴턴 방식은 m\*m개의 행렬을 모두 구하는 것이 아닌, 단위 행렬에 갱신 규칙을 적용해 계산하는 형태를 지닙니다.  **Q3**  앙상블 기법이 궁금합니다.  **A3**  여러개의 분류기(모델)을 생성하고, 그 예측을 결합함으로써 정확한 예측을 도출하는 기법이다. 하나의 강력한 모델 대신 약한 모델 여러개를 조합해 더 정확한 예측에 도움을 주는 방식이다.  부트스트래핑은 교체와 함께 무자구이 샘플링을 사용하는 모든 테스트 또는 메트릭이다. 즉 부트스트랩 샘플링을 하면 샘플 데이터셋을 n개의 샘플 데이터셋을 가지고 있는 효과를 누릴 수 있게 된다. 배깅은 부트스트래핑 샘플링을 이용하여 여러 모델을 훈련하고 각 모델의 예측 결과를 투표방식 또는 평균 방식으로 집계한다. 즉. 주어진 하나의 데이터로 학습된 모델보단 더 좋은 모델을 만들 수 있는 앙상블 기법이다. 부스팅은 분류하기 어려운 특정 교육 샘플에 중점을 두는 방법으로, 여러 개의 약한 학습모델이 순차적으로 학습-예측하는 과정에서 이전의 학습 모델의 잘못 예측한 데이터에 가중치를 부여함으로써 오류를 개선해 강력한 모델을 구축하는 방법이다.    즉. 배깅과 부스팅은 여러 개의 모델을 학습시킴으로써 하나의 모델에서는 얻을 수 없는 성능과 안정성을 이끌어 내어 최적화된 모델을 만든다고 생각하면 된다. 두 모델의 차이점은 여러개의 모델을 어떻게 학습시키고 학습된 모델을 예측에 어떻게 활용하는지에 대한 접근 방법 차이가 있다.  **Q4**  기계학습을 통해 블라인드 원음 분리 문제를 해결하는 다양한 방법에 대해 궁금합니다.  **A4**  1. 주성분 분석 (Principal Component Analysis, PCA):  - PCA는 입력 데이터의 주성분을 찾아 이를 기반으로 데이터를 변환하는 방법입니다.  - 주성분은 데이터의 분산을 가장 많이 설명하는 방향으로 정의됩니다.  - 블라인드 원음 분리에서는 혼합된 신호가 각 주성분에 어떻게 투영되는지를 분석하여 각 원음을 추출할 수 있습니다.  2. 독립 성분 분석 (Independent Component Analysis, ICA):  - ICA는 주어진 혼합 신호로부터 독립적인 원음을 분리하는 데 사용됩니다.  - 혼합된 신호가 통계적으로 독립적이라는 가정을 기반으로 합니다.  - ICA는 각 원음의 독립성을 최대화하고 상호 의존성을 최소화하는 선형 변환을 찾아냅니다.  3. 딥러닝 기반 방법:  - 딥러닝을 사용한 방법 중 하나는 Recurrent Neural Networks (RNNs)나 Long Short-Term Memory (LSTM)과 같은 순환 신경망을 활용하는 것입니다.  - 혼합된 신호의 동적인 특성을 캡처하여 원음을 추출할 수 있습니다.  - 또는 Convolutional Neural Networks (CNNs)를 사용하여 시간 및 주파수 특성을 학습할 수 있습니다.  4. 매스킹 기반 방법:  - 주파수 영역에서의 소스 신호에 해당하는 부분을 찾아내어 추출하는 방법도 있습니다.  - 예를 들어, 주파수 영역에서 소스 신호에 해당하는 부분을 강조하거나, 혼합 신호의 주파수 영역에서 소스 신호에 해당하는 부분을 제거하는 방식으로 작동할 수 있습니다.  5. 오토인코더 (Autoencoder):  - 오토인코더는 입력 데이터를 재구성하는 방식으로 학습됩니다.  - 원음 분리에서는 혼합된 신호를 입력으로 사용하고, 오토인코더는 각 원음을 재구성하도록 학습됩니다.  이러한 방법들은 각각 장단점이 있고, 문제에 따라 어떤 방법이 더 효과적인지 다를 수 있습니다. 선택한 방법은 주어진 데이터의 특성, 문제의 복잡성, 및 사용 가능한 계산 자원에 따라 달라질 수 있습니다.  **Q5**  켤레 그레이디언트 방법에 대해서 궁금합니다.  **A5**  켤레 그레이디언트 방법(Conjugate Gradient method)은 최적화 문제에서 함수의 최솟값을 찾는 방법 중 하나입니다. 특히, 제약이 없는 비선형 최적화 문제를 푸는 데 사용됩니다.  이 방법은 일반적인 경사 하강법과 비교하여 보다 효율적으로 최적화를 수행하는 방법 중 하나입니다. 켤레 그레이디언트 방법은 잘 조절된 단계 크기와 그레이디언트의 정보를 이용하여 최적점에 대한 탐색을 수행합니다  .  켤레 그레이디언트 방법은 다음과 같은 특징을 가지고 있습니다:  1. 수렴 속도: 일반적인 경사 하강법에 비해 수렴 속도가 빠릅니다.  2. 메모리 효율성: 이전 그레이디언트 방향에 대한 정보를 활용하여 새로운 방향을 찾는데 사용하여 메모리 사용량을 줄일 수 있습니다.  3. 비선형 최적화에 효과적: 비선형적인 함수의 최적화에 특히 유용합니다.  켤레 그레이디언트 방법의 주요 아이디어는 이전에 탐색한 방향으로 서로 켤레(conjugate)되는 방향으로 갱신하여 최적점을 찾아가는 것입니다. 이전 그레이디언트 방향으로는 이미 최소값을 찾을 수 있는 방향이기 때문에 새로운 방향은 그 이전의 방향과 켤레되도록 계산됩니다.  간단히 말하면, 켤레 그레이디언트 방법은 현재 위치에서의 그레이디언트와 이전에 탐색한 방향들의 조합을 이용하여 새로운 방향을 결정하고, 이를 통해 최적점을 찾아가는 방법입니다. 이 과정에서 반복적으로 함수의 값을 최소화하는 방향을 찾아가며 최적화를 수행합니다. | |
| 질문내용 | Q:  하이퍼 매개변수를 최적화할 때의 핵심은 하이퍼 매개변수의 ‘최적 값’이 존재하는 범위를 조금씩 줄어 나가는 것입니다.  이때, 하이퍼 매개변수의 범위는 대략적으로 지정하는 것이 효과적이라 로그 규모 간격을 사용한다 합니다. 그렇다면, 실제로 로그 규모 간격이 쓰이는 경우는 어떤 케이스가 있는지 궁금합니다. | |

**개인첨부레포트**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 학과 | 컴퓨터 전자시스템공학 | 학번 | 201904458 | 이름 | 이준용 |
| 구분 | 내용 | | | | |
| 학습 범위 | 기계학습 5장 딥러닝 최적화  5.3 규제의 필요성과 원리  5.4 규제 기법  5.5 하이퍼 매개변수 최적화  5.6 2차 미분을 이용한 최적화 | | | | |
| 학습 내용 | **기계학습 5장 딥 러닝 최적화**  **5.3 규제의 필요성과 원리**  과잉 적합에 빠지는 이유 -> 대부분 가지고 있는 데이터에 비해 훨씬 큰 용량의 모델을 사용하기 때문.  ➔ 훈련집합을 단순히 ‘암기’ 하는 형식이 되어버림.    현대 기계 학습은 충분히 큰 용량의 모델을 설계한 다음, 학습 과정에서 여러 규제 기법을 적용한다.  규제  모델 용량에 비해서 데이터가 부족한 경우의 불량 문제를 해결하기 위한 수단  적절한 가정을 투입하여 문제를 품  - 입력과 출력 사이의 매핑은 매끄럽다는 사전 지식.  -> ‘수치적 용량’을 제한하거나 비지도 학습을 이용.  - 오래 전부터 수학과 통계학에서 연구해온 주제.  Optical Flow : 인접한 Frame이 있을 때 Pixel의 움직임이 해당 개체를 어떻게 이동시키는가?  Nonlinear Image Registration : Deform을 시킬 때 해당 Field가 어떻게 변하느냐를 추적.    일반화 오류를 줄이려는 의도를 가지고 학습 알고리즘을 수정하는 방법 모두를 규제라고 할 수 있다.  **5.4 규제 기법**  \* 명시적 규제 : 가중치 감쇠나 드롭아웃처럼 목적함수나 신경망 구조를 직접 수정 방식  \* 암시적 규제 : 조기 멈춤, 데이터 증대, 잡음 추가, 앙상블처럼 간접적으로 영향을 끼치는 방식  가중치 벌칙  티호노프의 규제 기법 중 관련 변수가 드러나도록 할 경우에 해당  규제항은 훈련 집합과 무관하며 데이터 생성 과정에 내재한 사전 지식  매개변수를 작은 값으로 유지하게 하므로 모델의 용량을 제한 -> 수치적 용량 제한  MLP와 DMLP에 적용됨.  λR(θ)를 어떤 것으로, 얼마나 설정할 것인가? – 보통 L2놈이나 L1놈을 사용  \* 놈 : 벡터의 크기, 혹은 길이. L1 -> 절댓값의 합, L2 -> 자승의 합(유클리드 놈)    ∗ λ = 0으로 두면 규제를 적용하지 않은 원래 식 θ = θ - ρ∇J이 됨. -> 가중치 감쇠는 θ에 (1 − 2ρλ)를 곱해주기만  함.   * 최종 해를 원점 가까이 당기는 효과.   선형 회귀에 적용할 경우  공분산 행렬 𝑋𝑇𝑋의 대각 요소가 2λ만큼씩 증가  -> 역행렬을 곱하는 셈이 되므로 가중치를 축소하여 원점으로 당기는 효과  규제식 L1놈의 Gradient 계산    ➔ 부호만 고려한다   * L1놈의 희소성 효과 – 0이 되는 매개변수가 많음   -> 선형 회귀에 적용하면 특징을 선택하는 효과를 보여준다.  조기 멈춤  일정 시간(𝑇𝑎𝑝𝑡)이 지나면 과잉 적합 현상이 나타남 -> 일반화 능력이 저하 -> 훈련 데이터를 단순히 암기하는  문제가 발생  조기 멈춤은 검증집합의 오류가 최저인 점 𝑇𝑎𝑝𝑡에서 학습을 멈춘다.  실제 학습은 최저값 갱신이 한 번만 이루어 지지 않으므로 여러 최저값들의 parameter를 저장한 후 가장 낮은  최저값을 기록하는 parameter 𝑇𝑎𝑝𝑡를 선택하여 멈춘다.  데이터 확대  과잉적합을 방지하는 가장 확실한 방법으로, 큰 훈련집합을 사용한다.  -> 데이터 수집은 비용이 많이 든다는 한계가 있음.   * 데이터를 인위적으로 변형하여 확대함 : 자연계에서 벌어지는 잠재적인 변형을 프로그램으로 흉내   Ex : 모핑을 이용한 변형, 자연 현상 확대, 잡음 추가 등  드롭아웃  배치 정규화가 나오기 전까지 많이 사용된 방법으로, 입력층과 은닉층의 노드 중에서 임의로 선택하여 제거하는  방식   * 많은 부분 신경망을 자체적으로 제작하여 예측 단계에서 앙상블 결합 기법을 이용.   ➔ 계산 시간과 메모리 공간 측면 등, 비용에 대한 부담 문제가 발생.  앙상블  서로 다른 여러 개의 모델을 결합하여 일반화 오류를 줄인다.  서로 다른 예측기를 학습하는 일.  Ex : 배깅(훈련 집합을 여러 번 샘플링 하여 서로 다른 훈련집합을 구성), 부스팅(i번째 예측기의 오류를 i+1번째 에서 잘 인식하도록 의도적으로 구성)  이전에는 규제 기법으로 보지 않았으나 현대에 와서 규제에 대한 정의가 넓어짐에 따라 앙상블도 규제에 포함  **5.5 하이퍼 매개변수 최적화**  \* 내부 매개변수 : 에지 가중치로서 보통 θ로 표기, 학습 알고리즘이 최적화 함  \* 하이퍼 매개변수 : 모델의 외부에서 모델의 동작을 조정. Ex : 은닉층의 개수, CNN 마스크 크기와 보폭, 학습률 등.  하이퍼 매개변수 선택  표준 문헌이 제시하는 기본값을 사용   * 보통 여러 후보 값 또는 범위를 제시하며, 후보 값 중에서 주어진 데이터에 최적인 값을 선택함. * 하이퍼 매개변수 조합을 생성하는 방법에 따라 수동 탐색, 격자 탐색, 임의 탐색으로 나뉜다.   격자 탐색과 임의 탐색  ->보통 격자 탐색보다는 임의 탐색이 더 유리하다.  로그 규모 간격  어떤 매개변수는 로그의 규모를 사용해야 한다.  Ex : 학습률 볌위가 0.0001 ~ 0.1일 때   * 등간격 : 0.0001, 0.0002, 0.0003, …, 0.0998, 0.0999, 0.1 * 로그 규모 : 0.0001, 0.0002, 0.0004, 0.0008, …, 0.0256, 0.0512 …   매개변수가 m개이고 각각이 q개 구간이라면 𝑞  𝑚개의 점을 조사해야 함 -> 차원의 저주  **5.6 2차 미분을 이용한 방법**  \* 1차 미분을 이용한 방법 : 경사 하강법 – 현재 기계 학습의 주류 알고리즘  경사 하강법을 더 빠르게 할 수 있는가에 대해서 연구하다가 발견  뉴턴 방법  테일러 급수를 이용한 방법.  2차 함수에 뉴턴 방법을 적용했으므로 3차 항 이상을 무시한 식을 사용했음에도 최적의 경우를 제시함     * 기계 학습이 사용하는 목적 함수는 2차 함수보타 더 복잡한 형태이므로 한 번에 최적해에 도달하기는 힘   들다  ➔ 반복하는 뉴턴 방법을 사용해야 함.   * H를 구해야 함.   ➔ 매게 변수의 개수를 m이라 할 때 O(𝑚3  )라는 과다한 계산량 요구  ➔ 켤레 그레이디언트 방법이 대안으로 제시됨.  켤레 그레이디언트 방법  - 직선 탐색 방법  경사 하강법을 이용하여 얻은 해에 특정 값을 이용하여 해까지 일직선으로 도달하게 함  켤레 그레이디언트 방법의 식     * 𝑃𝑡와 𝑃𝑡−1를 켤레라고 부름.   유사 뉴턴 방법  H를 직접 구하는 대신 H의 역행렬을 근사하는 M을 사용한다.  처음에는 단위 행렬 I로 시작하여 그레이디언트 정보를 이용하여 점점 개선함.(LFGS)  기계 학습에서는 M을 저장하는 메모리를 적게 쓰는 L-BFGS를 주로 사용 | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 학과 | 컴퓨터전자시스템공학과 | 학번 | 201800615 | 이름 | 김동규 |
| 구분 | 내용 | | | | |
| 학습 범위 | 5장 딥러닝 최적화  5.3 규제의 필요성과 원리  5.4 규제 기법  5.5 하이퍼 매개변수 최적화  5.6 2차 미분을 이용한 최적화 | | | | |
| 학습 내용 | 과잉적합   * 모델 용량의 증가   + 일반적 데이터와의 차이가 점점 벌어짐   + 적당한 시점에서 끊어야 함   + 큰 용량의 구조로 먼저 계산한 후, 규제기법을 통해 필요한 데이터를 구하게 유도함 * 규제   + 모델 용량에 비해 데이터가 부족한 경우   + 특정 가정을 추가해 데이터를 추가함   + 원 목적함수 + 규제항으로 구성함      * + 원 목적함수 뿐만 아니라 규제항도 최소가 되는 지점을 찾아야 함   + 규제항은 훈련 집합에 의한 데이터와는 관련이 없음   + 용량 제한 및 데이터 생성의 사전지식과 관련이 있음   + 활용     - 특정 기준을 가지고 복잡도를 낮춤(티호노프의 규제 기법, 자연 상의 데이터들은 특정 규칙을 가지고 매끄러운 성질을 만족함) * 가중치 벌칙 * 가중치 감쇠   + L2놈     - 규제항이 추가됨으로서 매개변수의 갱신 식이 생겨남     - 해당 갱신 식을 계속 반복함으로서 최종해를 원점 가까이 당김     - 선형 회귀에서의 적용(n개의 요소를 가지는 고차원의 평면)       * X={x1,x2,x3…} = x개의 요소들       * W={w1,w2,w3…} = 각 x에 대한 가중치들       * Y={y1,y2,y3…} = 각 x와 w의 계산값에 대응하는 실제 측정값의 요소들       * 일반적으로는 x1w1+x2w2…=yi로 각 값에 대응하게 됨       * 모든 샘플(x)의 값들이 일정한 경우, Xw=y로 하나의 평면 위에 존재하게 됨(오차가 없음)       * 이때 가중치 감쇠를 적용하면 아래와 같이 나옴      * + - * 유클리드 거리(실제 직선거리)를 통해 구해짐(각 오차는 오차-결과의 제곱)   + L1 놈     - 규제항은 모든 개별요소의 절댓값의 합으로 표현됨     - 미분된 규제항이 사인항을 가지게 됨     - 맨해튼 거리(특정 축을 사용해서 해당 벡터를 표현함, 각 오차는 |오차|-|결괏값| * 조기멈춤   + 과적합이 일어나기 이전에 먼저 학습을 정지시킴   + 검증 집합의 오류율이 최저이면 정지함 * 데이터 확대   + 많은 양의 데이터를 사용해서 과적합 방지   + 데이터의 인위적 변형   + 모핑     - 원 데이터와 비슷하지만 비선형 변환을 통해 데이터 생성   + 노이즈 추가 * 드롭아웃   + 일부 입력층, 은닉층을 임의로 제거함   + 남은 입력층과 은닉층을 사용해서 결합함   + 앙상블 기법(여러 형태의 모델들을 사용해 각각의 모델에 맞는 특징을 추출한 후, 나중에 합치는 방법)을 활용해 데이터를 추출할 수 있음 * 하이퍼 매개변수 최적화   + 모델 외부에서 모델의 동작 조정   + 격자 탐색     - 각각의 매개변수를 일정한 간격으로 나눈 후, 교차점의 값 사용   + 임의 탐색     - 난수를 이용한 간격 사용 * 2차 미분을 이용한 방법   + 일반적인 최적화는 그레디언트를 활용함(1차)   + 대용량 최적화에서는 다른 방법을 활용함     - 그레디언트 잡음 줄이기       * 미니 배치     - 2차 미분의 활용       * 1차 미분(현재의 기울기)를 예측하는 2차 미분(기울기의 변화율)을 활용해 최적해를 파악함       * 뉴턴 방법         + 테일러 전개(다항식이 아닌 식을 다항식으로 표현하기)를 활용한 후, 식의 2차 미분까지 적용하면      * + - * + 이때 해당 q(x)의 최솟값을 구하기 위해 한번 더 미분하면         + 각 항이 0이 되어야 하므로, 함수의 관계가 나타남.         + 2차 미분값을 분모로 가지므로, 모든 기울기가 거의 0인 그래프에서는 오버슈팅이 발생할 수가 있음       * 켤레 그레디언트         + 주어진 값 중에서 최대한 해와 가까운 값을 추정함         + 기본 원리 = 현재 이동한 방향의 벡터(행렬)와 비슷한 방향으로 이동할 것이다 라고 추정(이전 벡터를 활용) | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 학과 | 철학과 | 학번 | 201802344 | 이름 | 유정훈 |
| 구분 | 내용 | | | | |
| 학습 범위 | 동영상 12주차 | | | | |
| 학습 내용 | 1. 과잉 적합에 빠지는 이유   (1)학습모델의 용량이 커지면서 일반화된 값과의 오류가 커짐.  (2) 대부분 가지고 있는 데이터에 비해 큰 용량의 모델 사용  (3)훈련집합을 단순히 암기하는 과잉적합에 주의 기울여야 함  (4)현대에는 학습과정에서 여러 규제 기법을 사용   1. 규제   (1)명시적 규제: 가중치 감쇠나 드롭아웃처럼 목적함수나 신경망구조를 직접 수정  (2)암시적 규제: 조기 멈춤, 데이터 증대, 잡음 추가, 앙상블처럼 간접적으로 영향을 미치는 방식   1. 가중치 벌칙   (1)규제항은 훈련집합과 무관하며 매개변수를 적은 값으로 유지하므로 모델의 용량 제한  규제항으로 L2놈이나 L1놈을 사용      (2)규제 항 R로 L2놈을 사용하는 규제 기법을 '가중치 감쇠(weight decay)라 부름  아래 역전파 과정에서 경사도를 계산하는 식이 아래 식    경사도를 경사도를 구하는 규제화된 경사도 진행양을 구하는 수식  람다가 0값으로 규제를 적용하지 않으면 원래식 쎼타 - 그래디언트가 됨   1. 2\*학습률\* 람다)를 원래 가중치에 곱해주는 형태   원래 진행해야 하는 쎼타값을 줄이고 그래디언트 만큼 이동하는 개념    원래 진행해야 하는 쎼타값을 줄이고 그래디언트 만큼 이동하는 개념    L2 norm 은 원형적인 제한사항이 있으므로 위와 같은 형태가 됨      절대값을 활용하는 L1놈 같은 경우 미분하면 사인함수가 나오게 됨    식을 정리하자면 매개변수 갱신은 사인값에 의해 정해지게 됨    가중치가 회전하는 형태  L1놈은 희소성 효과가 있음 가중치 w1에서는 0이되고 w2에서 1인 값으로 정해질 가능성 높음  0이 되는 매개변수가 많음   1. 조기멈춤 2. 훈련이 오랜 시간이 지나면 집하에선 오류율이 낮아지지만 테스트집합에선 오류율이 올라감 3. 일반화 능력이 저하되며 훈련데이터를 단순암기하기 시작 4. 조기멈춤은 검증집합의 오류가 최저일 때 학습 멈춤 5. 데이터 확대 6. 과잉적합을 방지하는 가장 확실한 방법은 큰 훈련집합을 사용하는 것을 기반으로 데이터를 수집 7. 데이터 수집 비용이 비용 많이 듦 8. 데이터를 인위적으로 변형하여 확대 | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 학과 | 철학과 | 학번 | 201803758 | 이름 | 탁성재 |
| 구분 | 내용 | | | | |
| 학습 범위 | 기계학습 13주차 동영상 | | | | |
| 학습 내용 | 1. 과잉적합에 빠지는 이유:  - (1) 모델 용량 증가로 인한 오류 증가: 모델의 용량이 증가하면 훈련 데이터에 대한 학습이 더욱 세세해지므로, 훈련 데이터에 과도하게 적합될 수 있고, 이는 새로운 데이터에 대한 일반화 능력을 저하시킬 수 있습니다.  - (2) 대량의 모델 사용: 모델의 용량이 훈련 데이터보다 큰 경우, 훈련 데이터를 완벽하게 학습할 수 있으며, 이는 테스트 데이터에 대한 일반화를 방해할 수 있습니다.  - (3) 훈련 데이터 암기: 훈련 데이터를 단순히 암기하는 과정에서 발생하는 과잉적합에 주의를 기울여야 합니다.  - (4) 규제 기법의 부족: 현대에는 학습 과정에서 여러 규제 기법을 사용하지 않으면 모델이 과도하게 복잡해져 과잉적합에 빠지기 쉽습니다.  2. 규제:  - (1) 명시적 규제: 가중치 감쇠나 드롭아웃과 같이 목적 함수나 신경망 구조를 직접 수정하여 규제를 적용하는 방식입니다.  - (2) 암시적 규제: 조기 멈춤, 데이터 증대, 잡음 추가, 앙상블과 같이 간접적으로 규제를 가하는 방식입니다.  3. 가중치 벌칙:  - (1) 규제항의 효과: 규제항은 훈련 데이터와 무관하며, 매개변수를 작은 값으로 유지하여 모델의 용량을 제한합니다.  - (2) L2 놈을 사용한 가중치 감쇠: L2 놈은 가중치의 제곱의 합에 비례하는 규제를 적용하는데, 이를 가중치 감쇠(weight decay)라고 부릅니다.  - 수식 설명: 경사하강법에서는 원래의 가중치 갱신식에 (1 - 2 \* 학습률 \* 람다)를 곱하여 가중치를 갱신합니다.  - L1 놈의 특성: L1 놈은 절대값을 사용하므로 가중치 갱신에 사인 함수가 영향을 미치게 되어, 가중치가 희소성을 가지는 효과가 있습니다.  4. 조기 멈춤:  - (1) 과적합 방지: 훈련이 오랜 시간이 지나면 훈련 데이터에 대한 오류는 감소하지만, 테스트 데이터에 대한 오류는 올라가기 시작합니다.  - (2) 일반화 능력 저하: 모델이 훈련 데이터를 단순 암기하기 시작하면 일반화 능력이 저하될 수 있습니다.  - (3) 조기 멈춤 기준: 검증 데이터의 오류가 최저일 때 훈련을 멈추는 기법입니다.  5. 데이터 확대:  - (1) 과적합 방지: 가장 효과적인 과적합 방지 방법 중 하나는 큰 훈련 데이터를 사용하는 것입니다.  - (2) 데이터 수집 비용: 데이터를 수집하는 비용이 많이 들기 때문에, 데이터 확대는 인위적으로 데이터를 변형하여 다양한 샘플을 생성하는 방법을 포함합니다. | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 학과 | 컴퓨터 전자시스템공학 | 학번 | 201702899 | 이름 | 이학빈 |
| 구분 | 내용 | | | | |
| 학습 범위 | 5.3 규제의 필요성과 원리  5.4 규제 기법  5.5 하이퍼 매개변수 최적화  5.6 2차 미분을 이용한 최적화 | | | | |
| 학습 내용 | -규제의 필요성과 원리  과잉적합 되는 것을 방지하고 일 반화 성능을 향상시키기 위해 사용된다. 과잉적합은 모델이 학습 데이터에 너무 맞춰져서 새로운 데이터에 대한 예측 성능이 떨어지는 현상을 말한다. 규제는 이러한 과적합을 제어하여 모델의 성능을 개선하는 방법 중 하나이다. 규제의 주요 목적은 다음과 같다:   1. 일반화 성능 향상: 모델이 새로운 데이터에 대해서도 잘 작동하도록 만들어준다. 2. 과잉적합 방지: 학습 데이터에만 지나치게 적합되는 것을 방지하여 모델의 성능을 향상시킨다.   규제의 주요 원리는 모델의 복잡성을 제어하고 가중치(weights)의 크기를 조절하여 모델이 데이터에 너무 맞춰지지 않도록 하는 것이다. 주로 사용되는 규제 기법으로는 다음과 같은 것들이 있다:   1. L1 규제 (Lasso 규제):   가중치(weight)의 절대값 합에 패널티를 부여하여, 불필요한 특성들의 가중치를 0으로 만들어 특성 선택(feature selection)을 유도한다. 이는 모델에서 중요하지 않은 특성을 제거하여 모델을 단순화하고 해석력을 높인다.   1. L2 규제 (Ridge 규제):   가중치의 제곱합에 패널티를 부여하여 가중치의 크기를 제한한다. 이를 통해 가중치를 줄이고 모델의 복잡성을 감소시킨다. Ridge 규제는 모든 특성을 유지하면서 가중치를 작게 만들어 중요한 특성에 더 집중할 수 있도록 한다.   1. Elastic Net 규제:   L1과 L2 규제를 혼합한 방법으로, 두 가지 규제의 장점을 취합하여 사용한다. Elastic Net은 L1 규제의 특성 선택 능력과 L2 규제의 안정성을 함께 제공한다.   1. Dropout:   신경망(neural network)에서 사용되는 규제 방법 중 하나로, 학습 중에 무작위로 일부 뉴런을 비활성화시켜 모델의 복잡성을 줄여 과적합을 방지한다. 이를 통해 서로 다른 부분집합의 네트워크를 학습하여 각각의 네트워크의 결과를 평균내어 모델을 강건하게 만든다.   1. 이러한 규제 기법들은 모델의 학습 과정에서 추가적인 제약을 부여하여 과적합을 줄이고 일반화 성능을 향상시킨다. 선택한 규제 기법은 데이터의 특성과 모델의 복잡성에 따라 적절히 조절되어야 한다.   -하이퍼 매개변수 최적화  머신 러닝 모델에서 모델 자체의 파라미터(weights)가 아닌, 모델을 훈련시키기 위한 하이퍼 파라미터들의 최적 값을 찾는 과정을 말한다. 하이퍼 파라미터는 모델의 구조와 학습 방법에 영향을 주는 파라미터로, 모델의 성능에 큰 영향을 미친다.  하이퍼 파라미터에는 학습률, 은닉층의 개수와 크기, 규제 강도, 배치 크기, 트리 기반 모델의 깊이 등이 있다. 이러한 하이퍼 파라미터들을 최적화하여 모델의 성능을 극대화하고 과적합을 줄이는 것이 중요하다.  하이퍼 파라미터 최적화를 위한 주요 방법과 기법은 다음과 같다:  그리드 탐색(Grid Search):  가능한 모든 하이퍼 파라미터 값의 조합을 시도하여 최적의 조합을 찾는 방법이다. 하지만 하이퍼 파라미터 조합이 많을 경우 계산 비용이 많이 들 수 있다.  랜덤 탐색(Random Search):  랜덤으로 하이퍼파라미터 값의 조합을 선택하여 탐색하는 방법이다. 그리드 탐색보다 계산 비용이 적게 들면서도 좋은 성능을 보이는 경우가 많다.  베이지안 최적화(Bayesian Optimization):  이전 단계의 결과를 활용하여 가장 가능성 있는 하이퍼 파라미터 값을 선택하는 방법이다. 이전에 시도한 값들을 바탕으로 모델을 평가하고, 다음 시도할 값의 후보를 선택함으로써 효율적으로 최적값을 찾을 수 있다.  진화 알고리즘(Evolutionary Algorithms):  유전 알고리즘과 같은 진화적 방법을 사용하여 하이퍼 파라미터 공간을 탐색하는 방법이다. 개체(population)들을 변형시켜 가장 좋은 하이퍼 파라미터 조합을 찾는다.  하이퍼 파라미터 최적화는 모델의 성능과 일반화 능력에 중요한 영향을 미치며, 이를 효율적으로 수행함으로써 모델의 성능을 향상시킬 수 있다. 주어진 문제와 데이터에 대해 적절한 최적화 기법을 선택하여 모델의 하이퍼 파라미터를 조정하는 것이 중요하다.  -2차 미분을 이용한 방법  2차 미분을 이용한 방법 중 하나는 뉴턴 방법 이며, 뉴턴 방법은 손실 함수의 최솟값을 찾기 위해 사용된다. 이 방법은 1차 도함수와 2차 도함수의 정보를 사용하여 최적화 과정을 수행한다.   1. 현재 위치에서의 기울기(1차 도함수)와 곡률(2차 도함수)을 활용하여 이동 방향 결정:  * 뉴턴 방법은 현재 위치에서의 기울기와 곡률을 동시에 고려하여 이동 방향과 크기를 결정한다.  1. 손실 함수의 최솟값을 찾기 위한 반복적인 업데이트:  * 현재 위치에서 2차 미분(도함수의 미분)을 이용하여 손실 함수의 최솟값을 예측하고, 해당 위치로 이동한다. * 이후 반복적으로 이동하며 손실 함수의 값을 최소화하는 방향으로 이동한다.   뉴턴 방법은 빠른 수렴 속도와 높은 정확도를 가지지만, 손실 함수가 복잡하고 곡률이 크게 변하는 경우에는 부정확할 수 있다. 또한, 손실 함수의 2차 미분을 계산하기 위해서는 추가적인 계산 비용이 필요하다.  2차 미분을 이용한 방법 중 하나인 뉴턴 방법은 일반적으로 고차원의 복잡한 손실 함수에는 적합하지 않을 수 있으며, 이를 극복하기 위해 제한된 메모리나 계산 자원을 사용하는 방법들이 개발되고 있다.  -켤레 그레이디언트 방법  최적화 문제에서 함수의 최솟값을 찾는 방법 중 하나이다. 특히, 제약이 없는 비선형 최적화 문제를 푸는 데 사용된다.  이 방법은 일반적인 경사 하강법과 비교하여 보다 효율적으로 최적화를 수행하는 방법 중 하나이다. 켤레 그레이디언트 방법은 잘 조절된 단계 크기와 그레이디언트의 정보를 이용하여 최적점에 대한 탐색을 수행한다. | | | | |