

- \* 풀이를 반드시 적을 것! (답만 적는 경우는 오답 처리함)
- \* 이름, 학번을 기입 하였는지 다시 한번 확인!

1. (5점) 단순선형회귀모델 (Simple linear regression model) 에서 아래의 식을 증명하시오. 단,

$$\hat{y}_i = \hat{w}_0 + \hat{w}_1 x_i \text{ 이고, } \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \text{ 이다.}$$

$$\sum_{i=1}^n (y_i - \hat{y}_i) (\hat{y}_i - \bar{y}) = 0$$

2. (6점) p2.RData 를 로딩하시오. p2 데이터 프레임에서 입력변수는  $x_1, x_2$  이고 출력변수는  $y$  라고 할 때, 학습모델 (예측모델) 을  $y = w_0 (x_1)^{w_1} (x_2)^{w_2}$  이라고 하자. p2 데이터 프레임에 대해 비선형회귀분석 (Nonlinear regression analysis)을 수행하고자 한다. 아래 물음에 답하시오.

- (1)  $SST, SSR, SSE$  를 구하시오. (3점)
- (2)  $(x_1, x_2) = (3, 62.5)$  일 때,  $y$  를 예측하시오. (3점)

3. (20점) p3.RData 를 로딩하시오. p3 데이터 프레임에서 입력변수는  $x_1, x_2, x_3, x_4$  이고 출력변수는  $y$  라고 할 때, 능형회귀분석 (Ridge regression)을 이용하여 학습모델을 만들고자 한다. 문제의 모든 기호는 강의자료 LNML1-5의 24 페이지를 따른다. 아래 물음에 답하시오.

(1) 기존의 경사하강법(Gradient descent) 알고리즘을 개선한 것 중 기계학습에서 많이 쓰이는 알고리즘이 적응형적률 알고리즘 (adaptive moment algorithm, 이하 Adam) 이다. Adam 알고리즘은 경사하강법 알고리즘 보다 상대적으로 학습속도가 빠른 것이 장점이다. 아래는 Adam 알고리즘의 의사코드 (pseudo code) 이며 여기서  $\odot$  은 원소별 곱셈 연산이다.

**Require:**  $\alpha$  : Stepsize  
**Require:**  $\beta_1, \beta_2 \in [0, 1)$  : Exponential decay rates for the moment estimates  
**Require:**  $C(\mathbf{w})$  : Cost function  
**Initialize:**  
 $\mathbf{w}$  : Initial parameter  
 $\mathbf{m} \leftarrow \mathbf{0}^\top$  // Initialize 1st moment vector  
 $\mathbf{v} \leftarrow \mathbf{0}^\top$  // Initialize 2nd moment vector  
 $t \leftarrow 0$  // Initialize timestep  
**while** termination condition is not met **do**  
 $t \leftarrow t + 1$   
 $\mathbf{g} \leftarrow \frac{\partial C(\mathbf{w})}{\partial \mathbf{w}}$  // Get gradient vector at timestep  $t$   
 $\mathbf{m} \leftarrow \beta_1 \cdot \mathbf{m} + (1 - \beta_1) \cdot \mathbf{g}$  // Update 1st moment vector  
 $\mathbf{v} \leftarrow \beta_2 \cdot \mathbf{v} + (1 - \beta_2) \cdot \mathbf{g} \odot \mathbf{g}$  // Update 2nd moment vector  
 $\hat{\mathbf{m}} \leftarrow \frac{\mathbf{m}}{1 - (\beta_1)^t}$  // Compute bias-corrected 1st moment vector  
 $\hat{\mathbf{v}} \leftarrow \frac{\mathbf{v}}{1 - (\beta_2)^t}$  // Compute bias-corrected 2nd moment vector  
 $\mathbf{w} \leftarrow \mathbf{w} - \alpha \cdot \frac{\hat{\mathbf{m}}}{\sqrt{\hat{\mathbf{v}}} + \delta}$  // Update parameter  
**end while**  
**return**  $\hat{\mathbf{w}} = \mathbf{w}$

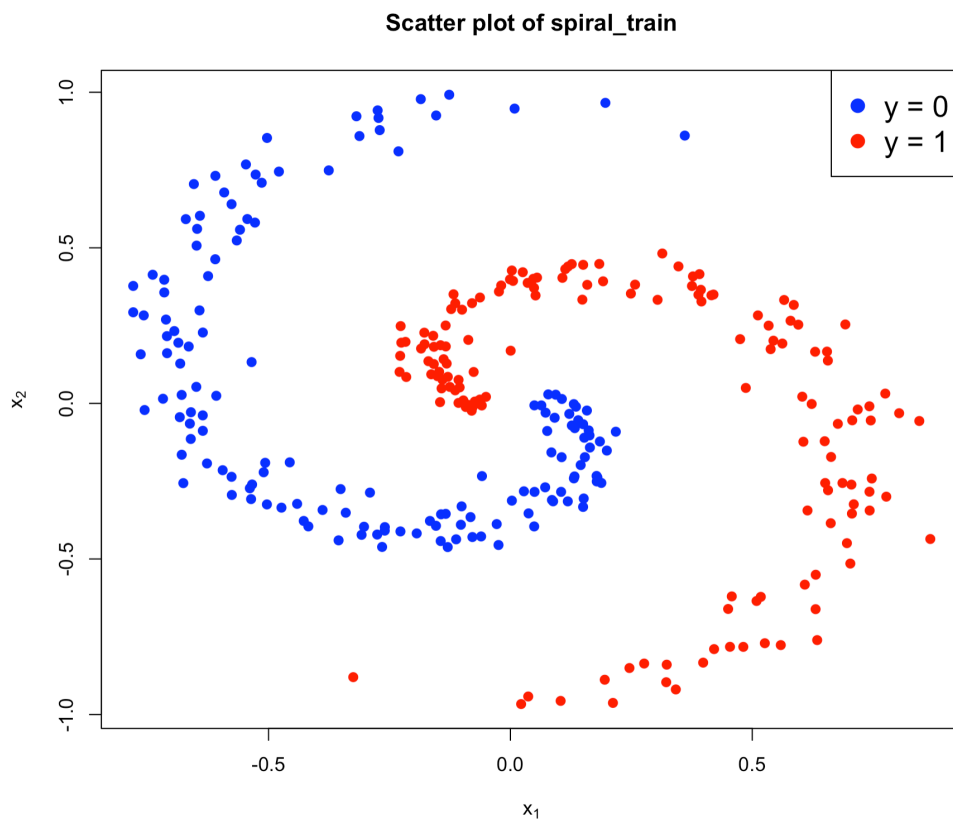
능형회귀모델의 비용함수가  $C(\mathbf{w}) = (\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda \mathbf{w}^\top \mathbf{w}$  일 때, Adam 알고리즘을 이용하여 학습모델의 파라미터 추정치  $\hat{\mathbf{w}}$  을 구하시오. 단,  $\alpha = 0.5$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\delta = 10^{-8}$ ,  $\lambda = 0.004$  로 설정하고,  $\mathbf{w} = (w_0, w_1, w_2, w_3, w_4)^\top = (0, 0, 0, 0, 0)^\top$  로 초기화 한다. 반복횟수는 19,000회 이다. (10점)

(2) Adam 알고리즘을 통해 추정된 능형회귀 학습모델의 결정계수  $R^2$  를 구하시오. (10점)

4. (17점) p4.RData 를 로딩하시오. p4\_train 데이터 프레임은 학습데이터(training dataset)이고, p4\_test 데이터 프레임은 검증데이터(testing dataset)이다. 입력변수는  $x_1, \dots, x_{100}$  이고 출력변수는  $y$  라고 할 때, 라쏘 회귀(Lasso regression)을 이용하여 학습모델을 만들고자 한다. 아래 물음에 답하시오.

- (1) 학습데이터를 이용하여 라쏘 회귀예측모델을 만들 때 출력변수  $y$  를 예측하는데에 사용되는 입력변수는 몇개인가? 단,  $\lambda = 0.04467$  라고 가정한다. (11점)
- (2) 검증데이터에 대해 예측오차제곱합의 평균을 구하시오. (6점)

5. (12점) p5.RData 를 로딩하시오. spiral\_train 은 학습데이터이고, spiral\_test 은 검증데이터이다. 입력변수는  $x_1$  과  $x_2$  이고 출력변수는  $y$  이다. 학습데이터에 대한 산점도는 아래와 같다. 로지스틱 회귀분석(Logistic regression)을 이용해 분류 예측모델을 구축하고자 한다. 아래 물음에 답하시오.



- (1) 입력변수 중  $x_2$  만 사용할 경우 검증데이터에 대해 정오표, 정확도, ROC 곡선 및 AUC를 구하라. (5점)
- (2) 입력변수를 모두 사용할 경우 검증데이터에 대해 정오표, 정확도, ROC 곡선 및 AUC를 구하라. (5점)
- (3) 입력변수를 모두 사용할 경우 과대적합(over-fitting)이 나타나는지 판단하시오. (2점)

6. (40점) 여름 방학동안 계획을 공백제외하고 290 ~ 300자로 서술하시오.

◎한 학기 동안 수고하셨습니다. 아래의 글귀를 항상 가슴에 담아두고 생활하시길 바랍니다.

一切唯心造(일체유심조) : 모든 것은 항상 마음가짐에 달려있다.