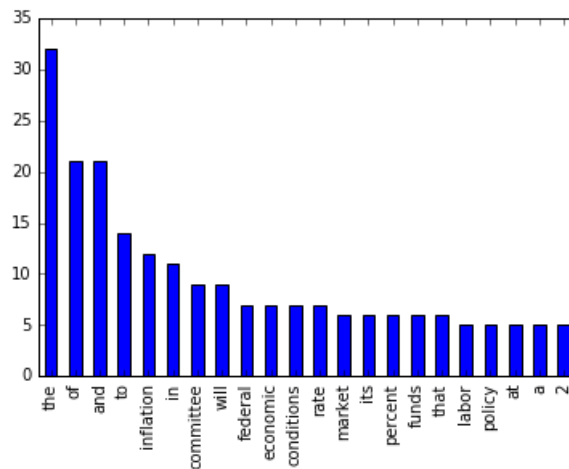


1. (20점) "data1.txt" 파일을 읽고 다음을 구현하시오.

- (1) 파일의 전문을 읽고 전체 단어의 개수를 구하시오. (단어는 앞뒤 공백으로 구분되는 단위를 의미함. 뒤에 붙은 쉼표, 마침표, 콜론, 세미콜론 등의 특수문자는 제거함) (10점)

of words = 499

- (2) 각 단어의 개수를 카운트해서 단어를 key로 하고 빈도를 value로 하는 dictionary 변수를 만드시오. 5회 이상 나온 단어만 뽑아서 개수를 내림차순으로 정렬하고 다음 그래프로 출력하시오. (단, 대문자와 소문자는 구별하지 않고 소문자로 변환하여 카운트함) (10점)



2. (15점) "고은택"은 보험사 A로부터 100년 동안 연말에 1억 원을 지급 받을 수 있는 연금을 가지고 있다. 그러나 보험사 A가 폐업하는 경우, 그 이후 지급되는 보험금은 받을 수가 없다. t년 경과 후 보험사 A가 생존해 있을 확률은 다음 식과 같다.

$$p(t) = e^{-\lambda t}$$

- (1) λ 가 인자로 주어질 때, 100년간 지급받을 보험금 합계액의 기대값을 구하는 함수를 구현하고, $\lambda = 0.01$ 일 때 기대값을 출력하시오. (Hint: 기대값은 다음의 식으로 계산할 수 있다) (10점)

$$\sum_{t=1}^{100} e^{-\lambda t}$$

Expectation = 62.897

- (2) 문제 (1)의 기대값이 80이 되기 위한 λ 의 값을 구하시오. (scipy.optimize 이용) (5점)

Lambda = 0.4592%

3. (15점) "씨니"는 서울 시내 30개의 치킨 체인점을 운영하고 있다. 30개 치킨 체인점 대표를 소집하는 회의 장소를 결정하려고 하는데, 각 점포로부터 직선거리의 합이 최소가 되는 지점을 선택하려고 한다. "data3.xlsx"는 각 체인점의 위치를 (X,Y) 평면에 표시한 좌표 값을 저장하고 있다.

- (1) 회의 장소를 (x, y)로 할 때, 총 거리를 계산하는 함수를 작성하고, 회의장소를 (10,10)으로 정할 때 총 거리를 출력하는 코드를 작성하시오. 단, (x, y)는 numpy의 array로 입력하시오. 반복문을 사용할 경우 50% 감점 (10점)

distance from (10,10) = 382.978

- (2) scipy.optimize 를 이용하여 직선거리의 합계가 최소가 되도록 하는 위치 (x, y)와 최소 거리를 구하시오. (5점)

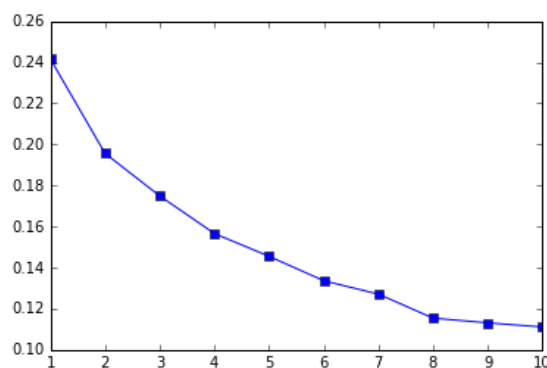
x=14.57 and y=11.87
Min distance = 360.06

4. (30점) 다음을 계산하는 몬테카를로 시뮬레이션 코드를 작성하시오.

- (1) "김신"과 "왕여"는 멀리뛰기 시합을 하는데, 각자 n 번을 뛰어서 가장 좋은 기록이 높은 사람이 이기기로 하였다. "김신"의 기록은 평균이 6m이고 표준편차가 1m인 정규분포이며, "왕여"는 평균이 7m이고 표준편차가 1m인 정규분포라고 할 때, "김신"이 이길 확률을 몬테카를로 시뮬레이션으로 계산하는 함수를 만들고 n=3일 때 확률을 출력하시오. 반복문을 사용할 경우 50% 감점 (시뮬레이션 회수는 10,000번으로 함) (10점)

Prob of Win = 0.168

- (2) 문제 (1)에서 n=1, 2, ..., 10일 때 각각 "김신"이 이길 확률을 계산하고 다음과 같은 그래프로 도시하시오. (5점)



- (3) "김신"과 "왕여"는 시합의 규칙을 다음과 같이 바꾸기로 하였다. 한 번씩 번갈아가며 뛰어서 뛴 거리의 합계가 먼저 40m를 넘는 사람이 이기기로 하고, 같은 회수를 뛰어서 동시에 넘거나 총 10번의 시행에도 둘 다 실패하면 비기기로 하였다. "김신"이 이길 확률과 비길 확률과 패할 확률을 각각 계산하시오. 반복문을 사용할 경우 50% 감점 (15점)

W: 1.08%, T: 19.14%, L: 79.78%

5. (20점) "data5.csv" 파일에는 미국채 2년, 10년 만기 수익률 (각각 CMT2, CMT10) 과 기대 인플레이션 (EINF) 의 시계열 데이터를 가지고 있다.

- (1) 데이터를 pandas.DataFrame 변수 x에 로딩하고, x의 index를 DATE로 변경하시오. (주의: 문자열로 로딩한 DATE 컬럼은 datetime 타입으로 변경해야 함) 결측치 (missing value)의 값을 직전 관측치로 채우고, 시계열 그래프를 다음과 같이 도시하시오. (10점)



- (2) CMT10과 CMT2의 차이를 이자율 term-spread 라고 한다. Term-spread를 daily로 계산하고, 매월별로 daily term-spread의 평균값을 계산하여 기대인플레이션(EINF)과 scatter plot으로 도시하시오. 단, 2010년 1월 부터 2016년 12월 까지 데이터만 이용하시오. (10점)

