

Praca domowa 4 – gauss

Szlachetka Miłosz
Nr albumu. 114014

1. Cel zadania

Celem zadania było sprawdzenie czy rozkład, otrzymanych przez servlet danych, jest rozkładem normalnym, a następnie wyznaczenie najmniejszej wartości poziomu istotności, dla której hipoteza o zgodności danych z rozkładem normalnym nie może być odrzucona. Poziom istotności należało wyznaczyć z dokładnością do dwóch miejsc dziesiętnych. Poziom istotności powinien zawierać się między wartościami 1 oraz 0.

2. Struktura programu

Program zawiera się w jednym pliku **Gauss.java**

-public class Gauss extends HttpServlet - klasa servletu, która realizuje wyżej wymieniony cel zadania

-protected void doGet(HttpServletRequest request, HttpServletResponse response) throws ServletException, IOException - obsługuje żądania HTTP GET. Metoda pobiera z adresu URL parametr n. Następnie dla pobranego parametru n zamieniane są przecinki “,” na kropki “.” tak, aby format zmiennoprzecinkowy był poprawny. Później następuje konwersja z typu String na double. Jeśli konwersja przebiegła niepoprawnie to następuje przerwanie obsługi żądania GET. Jeśli konwersja przebiegła poprawnie to z sesji (HttpSession) pobierany jest atrybut, który przechowuje listę dotychczas zebranych wartości (LinkedList<Double> - użyto tej struktury z tego względu, że wykonywane operacje wstawiania są na niej wyjątkowo szybkie. Szybsze niż na strukturze ArrayList). Jeśli takowa lista nie istnieje, to jest ona tworzona. Do listy dodawana jest wartość n typu double. Na końcu lista zapisywana jest w sesji (HttpSession).

-protected void doPost(HttpServletRequest request, HttpServletResponse response) throws ServletException, IOException - obsługuje żądania typu HTTP POST. Metoda pobiera z sesji (HttpSession) argument, który przechowuje listę wartości typu double (LinkedList<Double>). Jeśli taka lista istnieje to jest ona konwertowana do tablicy wartości double przy użyciu niżej opisanej metody convertDoubleListToDoubleArray. Jeśli utworzona

tablica posiada więcej niż dwie wartości double, to na jej podstawie tworzony jest rozkład normalny przy użyciu niżej opisanej metody `createNormalDistribution`. Następnie wyznaczany jest najmniejszy poziom istotności (alfa) przy użyciu metody `findSignificanceLevel` opisanej poniżej. Później otrzymana wartość alfa jest wypisywana z dokładnością do dwóch miejsc dziesiętnych, w taki sposób, że jest zaokrąglana w dół. Na końcu atrybut sesji przechowujący listę wartości Double (lista zgromadzonych z żądań GET wartości) jest usuwany, tak aby móc przeprowadzić proces szukania wartości alfa na nowo.

-private NormalDistribution createNormalDistribution(double[] data) - dla podanego parametru "data" wylicza średnią oraz odchylenie standardowe, a następnie tworzy obiekt reprezentujący rozkład normalny dla tych parametrów (średniej i odchylenia standardowego), a następnie zwraca go.

-private double findSignificanceLevel(double[] data, NormalDistribution normalDistribution) - wyznacza najmniejszą wartość poziomu istotności, dla której hipoteza o zgodności danych (podanych w argumencie "data") z rozkładem normalnym (podanym w argumencie "normalDistribution") nie może być odrzucona. W metodzie wykonywany jest Test Kołmogorowa-Smirnowa przy użyciu metody `kolmogorovSmirnovTest` z Apache Commons Math 3.4.1 API. Metoda ta przyjmuje tablicę wartości double oraz rozkład (w tym wypadku rozkład normalny). Dla tych argumentów sprawdza czy dane "data" pochodzą z rozkładu "normalDistribution" wyznaczając wartość p. Zgodnie z definicją, jeśli alfa (poziom istotności) jest mniejszy niż p-value to nie ma podstaw do odrzucenia hipotezy zerowej (czyli w tym wypadku nie ma podstaw do odrzucenia hipotezy o przynależności danych "data" do rozkładu normalnego). Zatem, zgodnie z warunkiem $\alpha < p\text{-value}$, od otrzymanej wartości p (p-value) odejmowana jest niewielka liczba ($\epsilon = 0.00000001$) i w ten sposób wyznaczana jest wartość alfa (poziom istotności). Jeśli alfa okaże się ujemna to metoda zwraca 0.

-private double[] convertDoubleListToDoubleArray(List<Double> doubleList) - metoda, która dla podanej (jako argument metody) listy wartości Double (`List<Double>`) konwertuje ją na tablicę wartości double (`double[]`) i zwraca ją (tablicę).

-private void clearSessionGauss(HttpSession session) - usuwa z sesji (`HttpSession`) atrybut, który przechowuje listę dotychczas zebranych (z żądań typu GET) wartości.

3.Mechanizmu poszukiwania rozwiązania

Podczas obsługi każdego żądania typu GET pobierana jest wartość parametru n z adresu URL. Wartość ta jest pobierana jako String, a następnie zamieniane są w niej znaki przecinka “,” na znaki kropki “.”. Po tej operacji następuje parsowanie do typu double. Potem z sesji (HttpSession) pobierany jest atrybut, który przechowuje listę dotychczas zgromadzonych wartości n (typu Double). Jeśli taka lista nie istnieje to jest tworzona. Do listy dołączana jest wartość n (typu double). Tak zmodyfikowana lista zapisywana jest jako atrybut sesji (aktualizacja atrybutu sesji). Ten sposób postępowania jest powtarzany dla każdego żądania GET.

Kiedy servlet otrzyma żądanie POST wykonywany jest następujący zestaw operacji. Z atrybutu sesji (HttpSession) pobierana jest lista zgromadzonych z żądań GET wartości typu Double. Jeśli taka lista istnieje (nie jest null) to jest ona konwertowana do tablicy wartości double (jest to wymagane, gdyż właśnie tego typu dane muszą zostać użyte podczas testu statystycznego w późniejszym etapie). Jeśli w tablicy wartości double znajduje się więcej niż jeden element to na podstawie tych danych (tablicy wartości double) tworzony jest rozkład normalny o średniej i odchyleniu standardowym wyznaczonym na podstawie danych w tablicy wartości double. Następnie dla zebranych danych (tablica wartości double) oraz utworzonego rozkładu normalnego, wykonywany jest test Test Kołmogorowa-Smirnowa, który działa w oparciu o wyznaczanie odległości dystrybuanty empirycznej od teoretycznej dla rozkładu normalnego. Klasa oraz metoda umożliwiająca wykonanie tego testu zawarta jest w Apache Commons Math 3.4.1 API. W wyniku tego testu otrzymywany jest parametr p-value. Zgodnie z definicją, jeśli alfa (poziom istotności) jest mniejszy niż p-value to nie ma podstaw do odrzucenia hipotezy zerowej(czyli w tym wypadku nie ma podstaw do odrzucenia hipotezy o przynależności danych do rozkładu normalnego). Dlatego też, od p-value odjęta zostaje niewielka liczba (epsilon = 0.00000001) i w ten sposób wyznaczana jest wartość alfa (poziom istotności). Otrzymany wynik (alfa) jest wypisywany z dokładnością do dwóch miejsc dziesiętnych. Dokładność ta jest uzyskiwana przy pomocy obiektu klasy “DecimalFormat”. Ustawiana jest liczba miejsc jedności na jeden oraz liczba miejsc po przecinku na dwa. Następnie ustawiane jest zaokrąglanie w dół i wykonywana jest metoda format, która zwraca String spełniające wcześniej określone wymagania. W otrzymanym stringu przecinki są zamieniane na kropki (zgodnie z formatem liczb double). Kod :

```
DecimalFormat df = new DecimalFormat("#.##");  
df.setRoundingMode(RoundingMode.FLOOR);  
df.format(significanceLevel).replace(",", ".")
```

,gdzie “significanceLevel” to wartość typu double o dowolnej liczbie miejsc po przecinku.