PEMULUSAN LOESS

#### Dr. Bagus Sartono, S.Si., M.Si. , Amalia Kholifatunnisa (G14170010)

#### sabtu, 08 Agustus 2020

Pendahuluan

Model regresi linear merupakan salah satu model yang sering digunakan dan disukai karena kesederhanaannya serta kemudahan dalam memahami hasilnya. Namun, banyak kasus yang terjadi di kehidupan ini tidak selalu menghasilkan bentuk linear yang sesuai. Untuk itu diperlukan metode yang mampu mengidentifikasi pola-pola tersembunyi dalam data. Salah satu metode pemulusan non-parametrik yang berkembang dan telah digunakan secara luas adalah *loess*(*locally estimated scatterplot smoothing*). Metode *loess* merupakan metode pemulus yang secara luas sering digunakan dengan sifat-sifat kekekaran (*robustness*) yang baik. Pemulusan *loess* bekerja dengan mencari penduga lokal bagi nilai peubah Y untuk beberapa nilai peubah X. *Loess* disebut sebagai running-line smoother terboboti, dan setiap garis lokal diduga mengunakan suatu metode robust menggantikan kuadrat terkecil. Sehingga *loess* menjadi pemulus yang kekar. Prosedur pengerjaan *loess* memerlukan komputasi yang intensif karena tahapan yang cukup banyak, namun proses perhitungannya relatif tidak sulit.

#### Pemulusan *Loess* (*Locally Estimated Scatterplot Smoothing*)

Pemulusan *Loess* lebih banyak dimanfaatkan untuk melihat hubungan antara dua peubah, meskipun telah ada pengembangan untuk regresi berganda. Gambaran umum terkait pemulusan *loess* yaitu misalkan terdapat gugus data dengan peubah X dan Y. Kedua peubah tersebut disajikan dalam plot tebaran dengan setiap titik pada plot adalah amatan \((x\_{i},y\_{i})\), untuk i = 1,2,…,n.

Prosedur *loess* dimulai dengan mementukan barisan m buah titik evaluasi \(X\_{j}\), j = 1,2,…,n. Tititk evaluasi ini merentang seluruh nilai peubah X dan berjarak sama. Kemudian,*Loess* akan menjalankan serangkaian m buah analisis regresi terboboti dengan melibatkan amatan \((x\_{i},y\_{i})\) yang berada di sekitar \(X\_{j}\). Hal ini disebut regresi yang bersifat “lokal” karena setiap analisisnya meilbatkan sebagian amatan yang berdekatan dengan titik evaluasi dilihat nilai peubah X-nya.

Analisis regresi lokal yang digunakan bisa berupa persamaan garis lurus, kuadratik, ataupun persamaan polinomial lain yang derajatnya dinyatakan dalam notasi \(\lambda\). Banyaknya amatan yang terlibat dalam analisis regresi ditentukan menggunakan acuan proporsi dari banyaknya data dan dinotasikan \(\alpha\). Sedangkan, untuk bobot yang digunakan pada analisis regresi lokal merupakan fungsi dari jarak antara amatan dengan titik evaluasi dilihat dari peubah X. Sehingga, amatan yang berdekatan \(X\_{j}\) akan diberi bobot besar dan memiliki pengaruh lebih besar dalam penentuan nilai dugaan lokalnya yang dinotasikan \(Y\_{j}\).

Selanjutnya, persamaan dan koefisien regresi yang diperoleh dari setiap model regresi lokal digunakan untuk menduga nilai \(Y\_{j}\) dengan cara mengevaluasi nilai \(X\_{j}\) kepada persamaan regresi yang terbentuk. Jika semua model regresi lokal telah diperoleh, maka kita akan memiliki m pasang nilai \(X\_{j},Y\_{j}\) yang kemudian diplotkan secara *overlay* dengan plot titik-titik amatan aslinya dan dihubungkan secara terurut hingga menghasilkan kurva pemulusan.

Pemulusan *loess* terkadang disebut juga sebagai *vertical sliding window* yang bergerak sepanjang sumbu horizontal pada plot tebaran dan berhenti di satu titik. Regresi yang dilakukan hanya melibatkan titik-titik yang ada dalam jendela, maka koefisien regresinya berubah mengikuti naik turunnya data. Hal ini menjadi keunggulana dari pemulusan *loess* karena mampu mengikuti pola rumit dan non-linear yang dibentuk oleh titik-titik amatan data.

Berikut ini tahapan pemulusan *loess* jika misalkan kita memiliki n buah amatan dan tiap amatan merupakan pasangan nilai dari kedua peubah yaitu \((x\_{i},y\_{i})\), i = 1,2,…,n.

1. Tahapan inisialisasi
   1. tentukan m buah titik yang selisihnya sama dan merentang seluruh nilai peubah X (sumbu horizontal). Titik-titik tersebut kita notasikan \(X\_{j}\) dengan j = 1,2,…,m.
   2. tentukan proporsi banyaknya data yang akan digunakan dalam pembentukan model regresi lokal, dinotasikan sebagai \(\alpha\). Sehingga,\(\alpha n\) menyatakan banyaknya amatan pada setiap model regresi.
   3. tentukan derajat polinomial model regresi lokal yang akan dibentuk \((\lambda)\), biasanya nilai yang digunakan yaitu 1 dan 2.
2. Tahapan penentuan titik pemulusan. Untuk setiap j = 1,2,…,n.
   1. hitung jarak setiap amatan \((x\_{i},y\_{i})\) ke titik evaluasi \(X\_{j}\) dengan formula \(d\_{i(j)}\) = \(|x\_{i}-X\_{j}|\).
   2. urutkan nilai \(d\_{i(j)}\) dari terkecil hingga terbesar. Misalkan jarak terurut dinotasikan \(d\_{[i](j)}\), kita ambil q = an titik dengan jarak yang paling kecil yaitu \(d\_{[1](j)}\), \(d\_{[2](j)}\), …,\(d\_{[q ](j)}\)
   3. hitung bobot untuk setiap q amatan terpilih sebagai fungsi dari jarak \[ w\_{i(j)} = f(d\_{i(j)})\]

fungsi penghitungan bobot antara lain adalah *tricube weight function* dalam bentuk \[ w\_{i(j)} = | 1- (\frac{d\_{i(j)}}{w\_{[q](j)}}) ^3 | ^3 \]lakukan pemodelan regresi lokal terboboti melibatkan q buah amatan terdekat dengan menduga koefisien model yang meminimumkan jumlah kuadrat galat terboboti

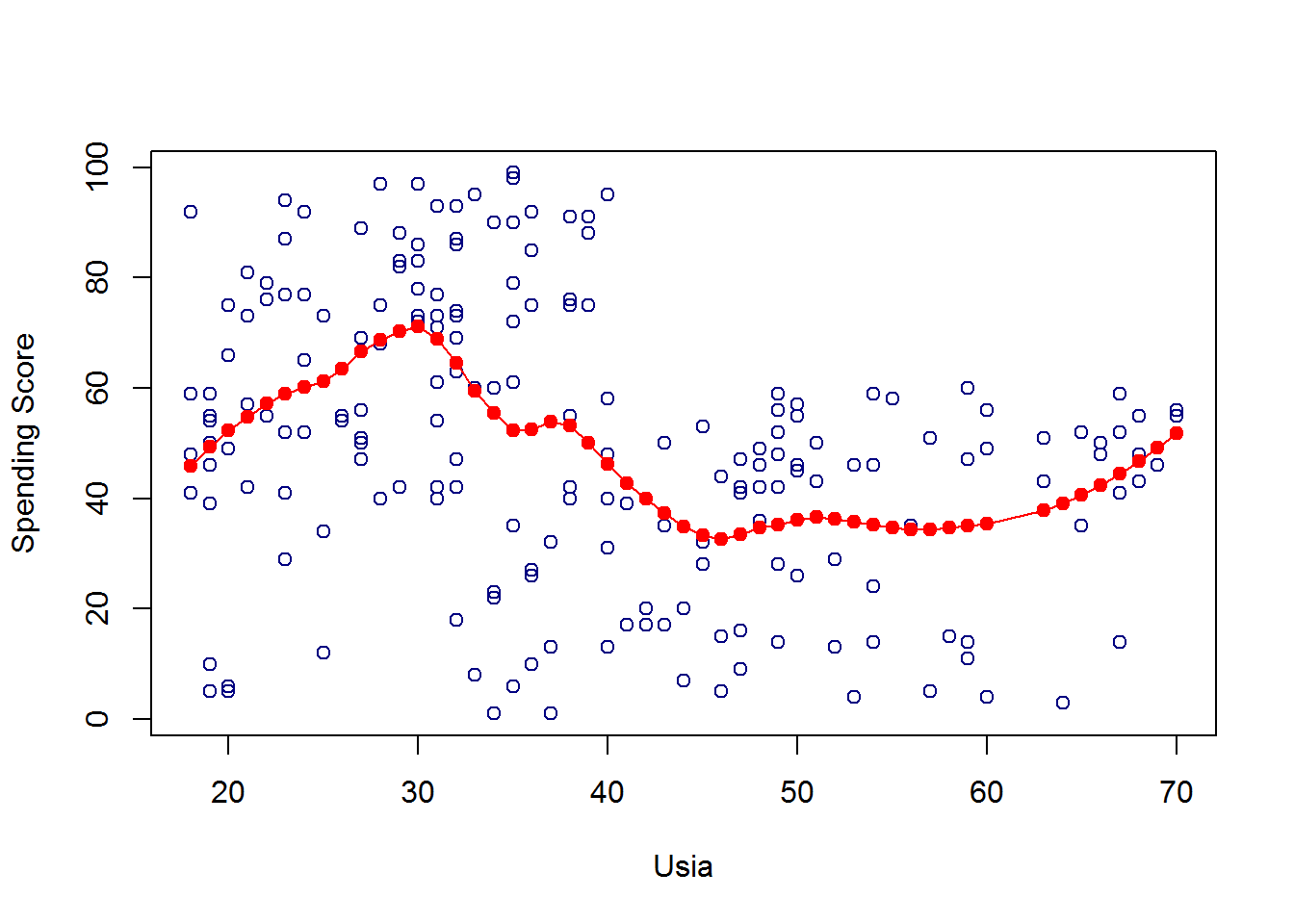
* 1. untuk \(\lambda\) = 1, modelnya linear, maka untuk mencari \(b\_{0}\) dan \(b\_{1}\) yang meminimumkan \[\sum\_{i = 1}^q w\_{i(j)}(y\_{i} - b\_{0} - b\_{1}x\_{i} ) ^ 2\]
  2. untuk \(\lambda\) = 2, modelnya linear, maka untuk mencari \(b\_{0}\), \(b\_{1}\), dan \(b\_{2}\) yang meminimumkan \[\sum\_{i = 1}^q w\_{i(j)}(y\_{i} - b\_{0} - b\_{1}x\_{i} - b\_{2}x\_{i}^2) ^ 2\]
  3. nilai \(w\_{i(j)}\) merupakan bobot yang diperoleh dari tahapan sebelumnya. lakukan penentuan nilai dugaan di titik evaluasi, yaitu \(Y\_{j}\) dengan formula \[Y\_{j} = \sum\_{k = 0}^1 b\_{k}x\_{i}^k\]

1. Tahapan penggambaran
   1. buat plot tebaran data \((x\_{i},y\_{j})\), i = 1,2,…,n.
   2. tambahkan titik \((X\_{j},Y\_{j})\), untuk semua j dan hubungkan titik-titik tersebut secara terurut berdasarkan nilai j.

Saat ini, kita telah dimudahkan oleh adanya program-program yang dikembangkan guna memudahkan proses penghitungan. Seperti halnya pemulusan *loess* yang dapat dikerjakan menggunakan *software* R dengan fungsi loess() pada packages stats(). Berikut ini ilustrasi pemulusan *loess* menggunakan dataset pengunjung mall dengan peubah penjelas usia dan peubah responnya *spending score*. Nilia \(\alpha\) yang digunakan sebesar 0.4 dan \(\lambda\) sebesar 2, artinya setiap model lokal dibangun melibatkan 40% amatan dan model yang digunakan model kuadratik.

mall <- read.csv("D:/DATASET AED/DATA/Mall\_Customers.csv", sep = ";")  
View(mall)

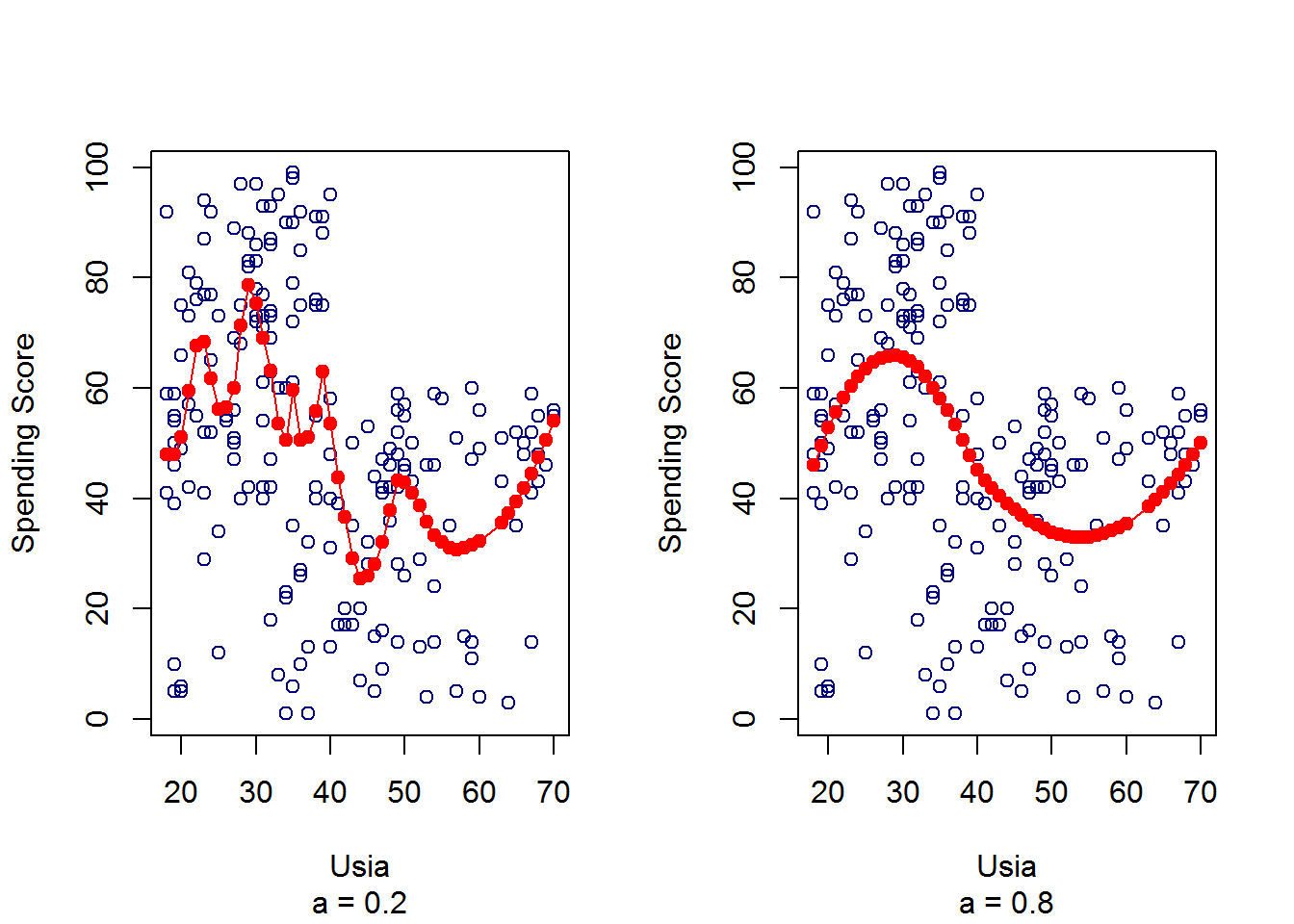
age <- mall$Age  
score <- mall$Spending.Score  
model\_loess <- loess(score~age, span = 0.40, degree = 2)  
pemulusan <- predict(model\_loess)  
hasil <- cbind(age,pemulusan) # nilai keluarannya berupa peubah usia dan hasil pemulusan  
hasil <- hasil[order(hasil[,1]),]  
plot(age,score, xlab = "Usia", ylab = "Spending Score", col ="navyblue")  
points(hasil, type ="o", col="red", pch =19, cex =1)



Berdasarkan syntax pemulusan *loess* opsi span berfungsi mengatur nilai \(\alpha\) dan opsi degree untuk mengatur nilai \(\lambda\). Kemudian pada syntax hasil <- cbind(age,pemulusan) menampilkan nilai pebah usia dan hasil pemulusannya secara berdampingan. Sedangkan untuk syntax hasil <- hasil[order(hasil[,1]),] berarti keluarannya berupa nilai peubah usia dan pemulusannya yang telah diurutkan.

Bedasarkan hasil pemulusan *loess*, lebar jendela menentukan tingkat kemulusan kurva gabungan dari titik-titik pemulusan. pada pemulusan *loess* lebar jendela diwakili oleh parameter \(\alpha\). Nilai \(\alpha\) yang semakin besar akan memberikan hasil yang lebih mulus. Untuk lebih jelasnya, berikut ini ilustrasi pemulusan *loess* dengan dataset pengunjung mall menggunakan \(\alpha\) = 0.2 dan \(\alpha\) = 0.8.

par(mfrow = c(1,2))  
  
#lebar jendela = 0.2  
age <- mall$Age  
score <- mall$Spending.Score  
model\_loess <- loess(score~age, span = 0.20, degree = 2)  
pemulusan <- predict(model\_loess)  
hasil <- cbind(age,pemulusan) # nilai keluarannya berupa peubah usia dan hasil pemulusan  
hasil <- hasil[order(hasil[,1]),]  
plot(age,score, xlab = "Usia", ylab = "Spending Score", col ="navyblue", sub = "α = 0.2")  
points(hasil, type ="o", col="red", pch =19, cex =1)  
  
#lebar jendela = 0.8  
age <- mall$Age  
score <- mall$Spending.Score  
model\_loess <- loess(score~age, span = 0.80, degree = 2)  
pemulusan <- predict(model\_loess)  
hasil <- cbind(age,pemulusan) # nilai keluarannya berupa peubah usia dan hasil pemulusan  
hasil <- hasil[order(hasil[,1]),]  
plot(age,score, xlab = "Usia", ylab = "Spending Score", col ="navyblue", sub = "α = 0.8")  
points(hasil, type ="o", col="red", pch =19, cex =1)



SEKIAN DAN TERIMA KASIH