

Dasar-dasar Pemrograman *Datamining* di R: *Plotting*

Sigit Wahyu Kartiko
gsigit[at]gmail[dot]com

Lisensi Dokumen:

Copyright © 2003-2006 IlmuKomputer.Com

Seluruh dokumen di IlmuKomputer.Com dapat digunakan, dimodifikasi dan disebarkan secara bebas untuk tujuan bukan komersial (nonprofit), dengan syarat tidak menghapus atau merubah atribut penulis dan pernyataan copyright yang disertakan dalam setiap dokumen. Tidak diperbolehkan melakukan penulisan ulang, kecuali mendapatkan ijin terlebih dahulu dari IlmuKomputer.Com.

Pendahuluan

Sebagian besar output yang dihasilkan dalam proses datamining disajikan dalam bentuk visual baik grafik maupun diagram. Penyajian informasi secara grafis membantu para pengambil keputusan untuk memicu *sense of analysis* melalui pola-pola visual tertentu. Oleh sebab itu dasar-dasar *plotting* bagi penyaji informasi menjadi poin yang cukup penting untuk dikuasai.

Kayanya *library* grafis di R merupakan salah satu keunggulan software statistik *opensource* ini. Kontribusi dari para developer R tersebut turut mempermudah para pengguna untuk memenuhi kebutuhan pelaporannya secara visual. Meski tidak secanggih produk komersial, tampilan grafis dari *library plotting* di R relatif cukup memadai untuk menghasilkan tampilan standar.

Dasar-dasar Plotting

Secara garis besar fungsi-fungsi membuat plot di R dibagi menjadi 3 bagian: perintah parameter layout plot, plot level tertinggi dan plot level terendah. Dengan demikian urutan dalam membuat plot adalah mulai dari penentuan parameter layout plot (jika diperlukan), plot level tertinggi baru kemudian menambahkan perintah plot level terendah (jika diperlukan).

Fungsi Plot Level Tertinggi (High-Level Plot)

Beberapa fungsi dasar pada level tertinggi antara lain:

| Fungsi | Deskripsi |
|---------|-------------------------------------------------|
| plot | scatter/line plot |
| hist | histogram |
| barplot | barplot |
| boxplot | boxplot |
| qqnorm | normal-quantile |
| qqplot | Q-Q plot |
| pairs | scatterplot matrix antara 2 atau lebih variabel |

Tabel 1: Fungsi plot level tertinggi

Untuk lebih jelasnya mari kita lanjutkan membuat plot berdasarkan dataframe yang sudah tersedia dalam package “datasets”.

```
> attach(cars)
> plot(speed, dist)
> plot(speed, dist, type="l") #titik-titik terhubung dengan garis (line)
> plot(speed, dist, type="b")
```

Tipe garis plot yang lain adalah sebagai berikut:

| Option | Plot type |
|----------|--------------------------------|
| type="p" | points (default)/ titik |
| type="l" | line / garis |
| type="b" | both line/pts (line dan titik) |
| type="n" | no (tanpa garis) |

Tabel 2: tipe-tipe garis plot

Selain garis plot, anda juga dapat menentukan warna, range dari x- dan y-axis, labels untuk x- and y-axis, dan sebagainya.

```
> plot(speed, dist, type="b", pch="x")
> plot(speed, dist, type="b", pch=5, col="red")
> plot(speed, dist, xlim=c(-1, 25), xlab="speed (mph)",
       ylab="distance, (ft)", ylim=c(-1, 150))
> plot(density(speed), type="l", main="normal, mu=0, sigma=1")
```

Fungsi plot memiliki beberapa pilihan dalam argumen fungsinya. Anda dapat menambahkan argumen di bawah dalam memkostumisasi tampilan plot.

| Option | Uraian |
|--------|--------------------------------|
| pch | plot character (pch=1, 2, ...) |
| lty | line type (lty=1, 2, ...) |
| lwd | line width (lwd= 1, 2,...) |
| col | color (col="red", "blue",...) |

| | |
|------|---------------------------------|
| xlim | x-axis limits: xlim=c(min,max) |
| ylim | y-axis limits |
| xlab | x-axis label: xlab="nama label" |
| ylab | y-axis label |
| main | judul utama |
| sub | sub judul |

Tabel 3: tipe-tipe garis plot

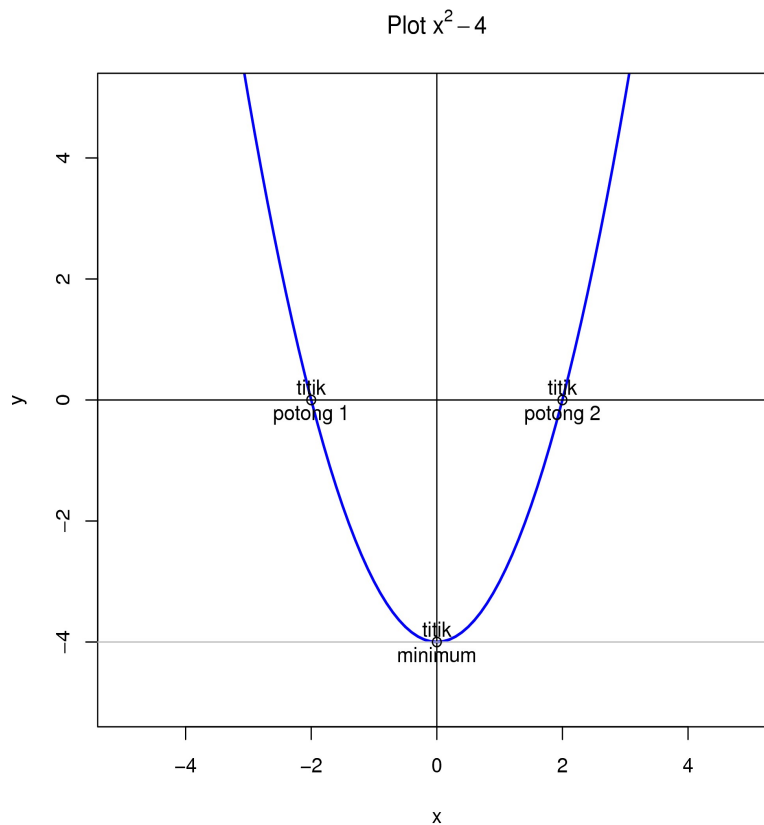
Untuk barplot, hist (histogram), boxplot dan qqnorm juga sebagian besar menggunakan opsi-opsi di atas dalam menampilkan grafik.

Fungsi Plot Level Terendah (Low-Level Plot)

Fungsi plot level terendah dapat diterapkan apabila fungsi plot tertinggi sudah dijalankan. Fungsi-fungsi plot level terendah dapat berupa informasi tambahan maupun lapisan grafik lain yang ditumpuk di atas plot level tertinggi. Sebagai contoh, setelah menjalankan fungsi plot anda dapat menambahkan keterangan titik potong tertentu, legenda, menambahkan gambar, menumpuk grafik lain sebagai sebagai pembanding.

Listing 1: Membuat kurva x^2-4 dan aksesorisnya

```
> #membuat plot fungsi tanpa data
> plot(function(x) x^2-4, ylim=c(-5, 5), xlim=c(-5, 5), ylab="y", xlab="x", col="blue", lwd=2)
> #menambah garis vertikal pada pada y=0
> abline(v=0)
> #menambah garis horizontal pada pada x=0
> abline(h=0)
> #menambah garis horizontal pada pada y=-4
> abline(h=-4, col="grey")
> #tanda bulat pada titik potong
> points(c(0, -2, 2), c(-4, 0, 0))
> #menambahkan keterangan pada titik-titik potong
> text(c(0, -2, 2), c(-4, 0, 0), c("titik\nminimum", "titik\npotong
1", "titik\npotong 2"))
> #memberi judul plot
> title(expression("Plot " * x^2 - 4))
```



Gambar 1: Kurva fungsi $f(x)=x^2-4$

Beberapa fungsi lain yang dapat di tambahkan dalam plot antara lain:

| Fungsi | Uraian |
|---------------|-------------------------------------|
| lines | menambah plot garis |
| points | menambahkan poin |
| abline | menambah garis |
| rect | menambah segi empat |
| text | menambah text |
| mtext | menambah text sesuai margin di plot |
| qqline | menambahkan baris pada qqnorm |
| title | menambah judul |
| arrow | gambar panah |

Tabel 4: tipe dekorasi yang dapat ditambahkan di plot

Fungsi Parameter Plot

Fungsi yang digunakan dalam membuat layout plot adalah `par()`. Fungsi ini dijalankan sebelum plot dibuat. Manfaat fungsi ini antara lain adalah untuk membuat beberapa grafik dalam 1 tampilan. Sebagai contoh apabila anda ingin menampilkan 4 plot dalam satu tampilan maka layoutnya adalah 2 baris x 2 kolom.

Jika urutan penambahan plot berdasarkan baris maka perintahnya adalah:

```
> par(mfrow=c(2,2))
```

Sedangkan jika penambahan plot berdasarkan kolom maka perintahnya adalah:

```
> par(mfcol=c(2,2))
```

Untuk mengembalikan tampilan dalam satu siklus perintah (tanpa menutup window plot secara manual) menjadi 1 plot penuh maka perintahnya adalah:

```
> mfrow=c(1,1)
```

Anda juga dapat mengubah warna tampilan default, karakter plot dsb. Namun opsi ini akan ditimpa ketika anda melakukan perubahan opsi di fungsi plot berikutnya setelah fungsi `par()`.

Untuk lebih jelasnya anda dapat melihat semua list fungsi plot dengan perintah berikut ini:

```
> library(help="graphics")
```

Praktek Membuat Plot

Membuat Kurva Biaya

Latihan Soal 1: Diketahui fungsi Total Biaya adalah:

$$C(q) = 4q^2 + 16q$$

Gambarkanlah kurva AC (Average Cost), MC (Marginal Cost), AVC (Average Variable Cost) dan nilai minimum dari AC!

Dalam teori mikroekonomi, biaya dalam perusahaan dibagi menjadi:

1. Biaya Total (*Total Cost*)

Biaya total adalah fungsi dari kuantitas (q) dengan satuan unit. Sebagai ilustrasi sederhana, jika suatu persamaan matematika biaya berupa persamaan kuadrat sbb:

$$TC(q) = aq^2 + bq + c$$

maka Biaya Total terdiri dari Biaya Variabel dan Biaya Tetap. Jadi secara matematis dapat juga di tulis sbb:

$$TC(q) = VC(q) + FC(q)$$

2. Biaya Total Rata-Rata (Total Average Cost atau Average Cost)

Biaya total rata-rata adalah biaya total per unit. Secara matematis, biaya total rata-rata diperoleh dari fungsi total biaya dibagi dengan kuantitasnya.

$$AC(q) = TC(q)/q$$

3. Biaya Tetap (*Fixed Cost*)

Biaya tetap adalah biaya yang tidak berubah pada saat unit bertambah. Dengan demikian dengan melihat persamaan total biaya maka biaya tetap merupakan konstanta yaitu c .

$$FC(q) = c$$

4. Biaya Variabel (*Variable Cost*)

Biaya Variabel adalah biaya yang meningkat seiring dengan bertambahnya unit yang diproduksi. Berdasarkan persamaan diatas maka fungsi biaya variabel adalah sbb:

$$VC(q) = aq^2 + bq$$

5. Biaya Variabel Rata-Rata (*Average Variable Cost*)

Biaya variabel rata-rata adalah biaya variabel per unit. Secara matematis, biaya variabel per unit diperoleh dari fungsi biaya variabel dibagi dengan kuantitasnya.

$$AVC(q) = VC(q)/q$$

6. Biaya Marginal (*Marginal Cost*)

Biaya marginal adalah setiap tingkat perubahan biaya yang harus dikeluarkan disetiap pertambahan unit. Secara matematis, maka biaya marginal adalah turunan pertama dari fungsi total biaya. Dengan demikian berdasarkan persamaan diatas maka fungsi biaya marginal adalah sbb

$$TC(q)' = MC(q) = 2aq + b$$

Untuk menjawab pertanyaan tersebut maka yang harus dibuat adalah fungsi-fungsi TC, AVC dan MC. Pertama kali yang dibuat adalah membuat ekspresi 3 fungsi tersebut.

1. Membuat fungsi TC, AC, AVC dan MC

Listing 2: Membuat kurva AC, MC dan AVC

```
> #membuat ekspresi fungsi TC, AC, AVC
> exp_TC <- expression(4*q^2+16)
> exp_AC <- expression((4*q^2+16)/q)
> exp_AVC <- expression((4*q^2)/q)

> #membuat ekspresi fungsi MC
> #untuk menghindari kesalahan penurunan maka menggunakan fungsi D()
> #(derivatif/turunan)
> exp_MC <- D(exp_TC, "q")

> #membuat ekspresi fungsi TC, AC, AVC dengan parameter q
> #eval() ditujukan untuk mengevaluasi ekspresi menjadi perintah
> f_AC <-function(q) {eval(exp_AC)}
> f_AVC <-function(q) {eval(exp_AVC)}
> f_MC <-function(q) {eval(exp_MC)}
```

2. Mencari nilai minimum AC.

Nilai minimum AC selalu berpotongan dengan fungsi MC, jadi nilai minimum AC

terjadi pada saat $AC=MC$. Nilai minimum/maksimum dapat dicari dengan fungsi `optimize()`. Tambahkan parameter `maximum=F` untuk menunjukkan bahwa yang dicari adalah nilai minimum. `c(0,10)` adalah vector rentang pengujian nilai q antara 0 s.d 10. Jika anda kurang yakin bahwa nilai q minimum adalah antara 0 s.d 10, saran penulis adalah gunakan rentang yang lebih besar misalnya `c(-100,100)`.

```
> #mencari nilai minimum AC
> min_AC <- optimize(f_AC, c(0,10), maximum=F)

> #menyimpan nilai minimum AC
> xmin<- min_AC$minimum
> ymin<- min_AC$objective
```

3. Membuat plot AC, AVC dan MC

Listing 3: Membuat kurva AC, MC dan AVC (lanjutan)

```
> #membuat rentang skala sumbu x dan y supaya dapat digunakan untuk
> #setiap plot
> y_lim <-c(0,40)
> x_lim <-c(0,6)

> #membuat kurva dari fungsi AC (high-level function)
> plot(f_AC ,xlim=x_lim, ylim=y_lim, col="orange", lwd=4,
      ylab="C/unit", xlab="q (quantity)")

> #menumpuk kurva dari fungsi MC diatas kurva AC (low-level function)
> #menambahkan parameter add=TRUE untuk menumpuk plot ini
> #parameter color dapat menggunakan kode warna HTML
> plot(f_MC, xlim=x_lim, ylim=y_lim, col="#FFFF00", lwd=4,
      add=TRUE)

> #menumpuk kurva dari fungsi AVC diatas kurva AC dan MC
> plot(f_AVC, xlim=x_lim, ylim=y_lim, col="brown", lwd=4,
      add=TRUE)
```

4. Memberikan keterangan dan asesoris lainnya pada grafik (*low-level function*)

a. Teks rumus di tempelkan pada kurva

Teks rumus di tempelkan pada kurva menggunakan fungsi `text()`. Parameter pertama merupakan kumpulan koordinat sumbu horizontal, dan parameter kedua adalah kumpulan koordinat sumbu vertikal. Sedangkan paramter ke-3 adalah kumpulan ekspresi text yang akan ditempelkan. Ekpresi tersebut menggunakan ekspresi yang biasa digunakan dalam format latex. Jika anda ingin mengeksplorasi notasi matematika maka anda dapat melihat contoh dengan mengetikkan `demo(plotmath)`

Listing 4: Membuat kurva AC, MC dan AVC (lanjutan)

```
> #memberi keterangan jenis kurva di koordinat tertentu
> text(c(4,4,4), c(f_AC(4)+1, f_MC(4)+1, f_AVC(4)-1),
      c(
        expression("AC ="* over((4*q^2+16), q)),
        expression("MC ="*8*q), expression("AVC ="*
        over((4*q^2), q))
```

```
)  
)
```

- b. Menghias nilai minimum AC dan perpotongan AC=MC

Nilai minimum AC akan ditunjukkan dengan menggunakan garis putus-putus dari sumbu horizontal maupun vertikal dengan menggunakan fungsi `rect()`. Teks koordinat ditempelkan dengan memasukkan variabel nilai minimum yang sudah dideklarasikan sebelumnya dengan `ymin` dan `xmin`. Menambahkan tanda panah berikut keterangan tanda panahnya.

Listing 5: Membuat kurva AC, MC dan AVC (lanjutan)

```
> #membuat garis putus-putus yang menunjukkan koordinat minimum AC  
> rect(0, ymin, xmin, 0,lty=2)  
  
> #menampilkan text koordinat nilai minimum  
> text(  
      c(xmin),  
      c(ymin)+1,  
      c(paste("MC =  
AC\n[" , round(xmin, 2) , " , round(ymin, 2) , "]" ) ) ,  
      cex=1.5)  
  
> #menggambar tanda panah untuk menunjuk titik minimum  
> arrows(1.5, 10, xmin, ymin,length=0.2,col="red")  
> #memberi keterangan tanda panah  
> text(1.5,10,c("titik potong\ndan\nminimum AC"))
```

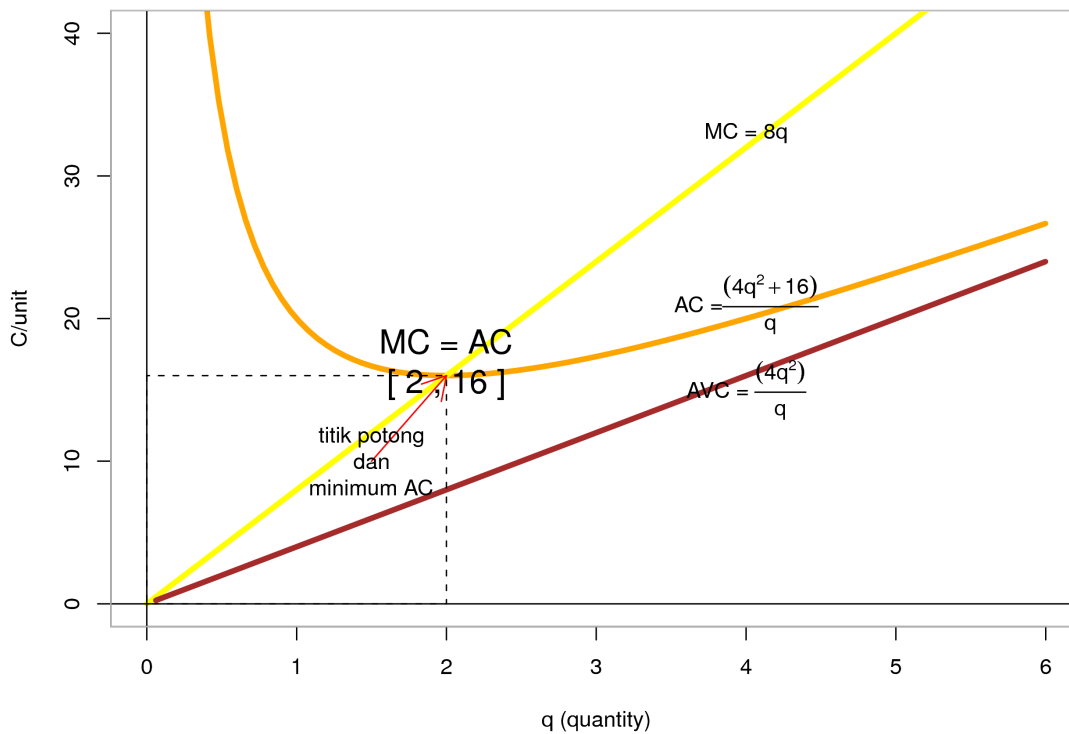
- c. Memberikan judul pada plot, menampilkan sumbu horizontal dan vertikal dengan garis tebal, dan menyamakan border plot.

Listing 6: Membuat kurva AC, MC dan AVC (selesai)

```
> #memberi judul pada plot  
> title(paste("Grafik AC (average cost),  
              MC (marginal cost), dan AVC (average  
variable cost)\nfungsi C(q)=", expression(4*q^2+16)))  
> #memberikan warna abu-abu pada sumbu koordinat  
> abline(h=0)  
> abline(v=0)  
> box(col="grey")
```

Hasil akhir dari kode program di atas adalah berupa grafik sebagai berikut:

**Grafik AC (average cost), MC (marginal cost), dan AVC (average variable cost)
 fungsi $C(q) = 4 * q^2 + 16$**



Penutup

Pembahasan di atas adalah dasar-dasar plotting di R sebagai prasyarat menuju pembahasan *datamining*. Dengan memanfaatkan fleksibilitas fungsi dasar tersebut diharapkan dapat mengoptimalkan penyajian visual melalui plotting, Setelah memahami plotting, materi yang berikutnya adalah pemrograman interface untuk menghubungkan fungsionalitas R dengan salah satu bahasa yang populer dalam pemrograman yaitu “java”.

Referensi

1. Ihaka, R. & Gentleman, R. (1996). *“R: A Language for Data Analysis and Graphics”*. *Journal of Computational and Graphical Statistics* 5 (3): 299–314. www.jstor.org
2. Rossiter, D. G. (2009). *Introduction to the R Project for Statistical Computing for use at ITC*. Accessed:03-03- 2010. [http:// www.itc.nl/ personal/ rossiter](http://www.itc.nl/personal/rossiter)
3. Williams, Graham J. *Rattle: A Data Mining GUI for R*. Accessed:06-04-2012. http://journal.r-project.org/archive/2009-2/RJournal_2009-2_Williams.pdf
4. Chihara, Laura. *Plotting Concepts in R*. Accessed: 07-12-2009. <http://www.people.carleton.edu/~lchihara/Splus/RPlot.pdf>
5. <http://www.R-project.org>

Biografi Penulis



Sigit Wahyu Kartiko. Menyelesaikan DIV di Sekolah Tinggi Akuntansi Negara dan meraih gelar Magister Ekonomi di Fakultas Ekonomi Universitas Indonesia. Memiliki kompetensi pada bidang *programming* (java, scala, R, C, C++), *database*, *accounting* (privat, publik), *public sector economics*.