

Data Mining - Latihan 2

Ali Akbar Septiandri
Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Al-Azhar Indonesia

June 2, 2017

Model Linear

1. Jelaskan dua pendekatan optimasi yang dapat dilakukan pada kasus regresi linear! *Petunjuk: Apa bedanya antara metode optimasi numerik dan analitis?*
2. Dalam kasus regresi linear, untuk $n \rightarrow \infty$ dan n adalah jumlah epoch, apakah kedua pendekatan optimasi tersebut akan berakhir pada titik optimum residual yang sama? Mengapa?
3. Gambarkan metode optimasi dengan *gradient descent* untuk kasus optimasi *convex*! Tunjukkan dalam gambar tersebut efek besarnya laju pembelajaran (*learning rate*; η)!
4. Jelaskan konsep *underfitting* dan *overfitting* pada tugas regresi! Anda dapat menggunakan gambar untuk membantu penjelasan Anda.

k-Nearest Neighbours

1. Apa yang terjadi saat kita memilih nilai $k = n$, dengan n adalah jumlah data?
2. Bagaimana cara memilih nilai k yang benar?
3. Apa yang dapat dilakukan untuk mengatasi perbedaan rentang variabel dalam klasifikasi dengan k-NN?
4. Diberikan data dengan dua atribut sebagai berikut:
A: [1, 5] B: [5, 1] C: [5, 5] D: [7, 3] E: [7, 9] F: [9, 7]
dengan A, C, E adalah kelas positif dan D, E, F adalah kelas negatif.
 - (a) Asumsikan bahwa Anda akan menghitung jaraknya dengan Euclidean *distance* dan $k = 1$. Gambarkan *decision boundary* yang akan dihasilkan oleh algoritma k-NN!
 - (b) Anda menerima data baru G: [7, 5]. Apa hasil klasifikasi untuk data tersebut?