58053-7 Algoritmien suunnittelu ja analyysi (kevät 2004)
Harjoitus 5 (26.–27. helmikuuta)

1. Osoita (esim. arvausta sovittamalla), että deterministisen select-algoritmin aikavaatimukselle T saadusta rekursioyhtälöstä

$$T(n) \leq T(\lceil n/5 \rceil) + T(\lfloor 3n/4 \rfloor) + cn$$

todella seuraa $T(n) = O(n)$.

2. Oletetaan annetuksi alirutiini, joka palauttaa joukon S *mediaanin* ajassa $O(|S|)$. Esitä tätä alirutiinia käyttävä yksinkertainen algoritmi, joka *mielivaltaisella* k löytää joukosta S sen k. suurimman alkion ajassa $O(|S|)$.

3. *Lähimmän parin ongelmassa* on annettu joukko $Q = \{ p_1, p_2, \ldots, p_n \}$, jossa on n erillistä tason pistettä $p_i = (x_i, y_i) \in \mathbf{R}^2$. Tehtävänä on löytää lähimpänä toisiaan olevat kaksi joukon pistettä, s.o. sellaiset $p_i, p_j \in Q$, $i \neq j$, että euklidinen etäisyys $d(p_i, p_j)$ on pienin mahdollinen.

   Esitä ongelmalle osittava algoritmi, joka on aikavaatimukseltaan $o(n^2)$ (eli asymptoottisesti tehokkaampi kuin triviaali $\Theta(n^2)$-ratkaisu).

   (*Vihje:* Cormen et al., luku 33.4.)

4. Tunnetun *Stirlingin kaavan* mukaan on $n! \approx \sqrt{2\pi n}(n/e)^n$. (Tarkemmin sanoen on $n! = \sqrt{2\pi n}(n/e)^n(1 + \Theta(1/n))$). Arvioi tätä kaavaa käyttäen binomikertoimen $\binom{n}{k}$ kertaluokkaa, kun (a) k on vakio, (b) n on parillinen ja $k = n/2$.

5. Koneoppimiseen liittyvässä *parhaan osavälin ongelmassa* on annettu n eri reaalilukua $x_1, x_2, \ldots, x_n$. Pisteet ovat väritettyjä: jokainen piste on joko valkoinen tai musta.

   Tarkastellaan osavälejä $I = [a, b]$, missä $a, b \in \mathbf{R}$. Osavälin I *virhe* annetulla pistejoukolla on

$$v(I) = |\, \{ x_i \mid x_i \text{ on valkoinen ja } x_i \notin I \} + |\, \{ x_i \mid x_i \text{ on musta ja } x_i \in I \} \,|,$$

   eli se on peittämättä jätettyjen valkoisten pisteiden lukumäärä plus peitettyjen mustien pisteiden lukumäärä. (Esim. kun peitetään kaikki valkoiset pisteet eikä yhtään mustaa pistettä, virhe on nolla.)

   Tehtävänä on löytää virheeltään pienin mahdollinen osaväli. Oletetaan pisteet annetuksi suuruusjärjestyksessä: $x_1 < \ldots < x_n$. Esitä ongelmalle taulukointiin perustuva algoritmi, joka tällöin löytää parhaan ratkaisun ajassa $O(n)$.