

E-M algoritmus pro odhad parametrů Gaussovské směsi

Předmět zadání

Proveďte odhad parametrů Gaussovské směsi $p(\vec{x}|\vec{\lambda}) = \sum_{m=1}^M c_m N(\vec{x}|\vec{\mu}_m, C_m)$ o neznámém počtu složek M , kde $\vec{\mu}_m$ je vektor středních hodnot m -té složky, C_m je kovarianční matice m -té složky, $\vec{\lambda}$ je vektor všech parametrů směsi a \vec{x} je posloupnost pozorování (2 rozměrných vektorů). Odhad proveďte algoritmem očekávání - maximalizace.

- Posloupnost pozorování \vec{x} načtete ze souboru `sp4_data.mat`, případně ze souboru `sp4_data.csv`, pokud vám MATLAB nevyhovuje (každý řádek obsahuje obě čárkou oddělené složky pozorování).
- Implementujte E-M algoritmus. **Zdůvodněte případné použití zjednodušujících předpokladů** (počet složek směsi, tvar kovarianční matice, atd.).
- Proveďte implementovaný algoritmus nad poskytnutými daty. Uvažujte zastavovací podmínku euklidovské vzdálenosti předcházejících a nově odhadnutých parametrů $\|\vec{\lambda}_i - \vec{\lambda}_{i-1}\| < 10^{-3}$, kde i je číslo iterace algoritmu.
- Tabelujte **všechny** hodnoty $i, \vec{\lambda}_i, \|\vec{\lambda}_i - \vec{\lambda}_{i-1}\|$.
- Do grafu vynesete závislost $\|\vec{\lambda}_i - \vec{\lambda}_{i-1}\|$ na počtu iterací.

Použité nástroje

Simulaci proveďte v prostředí MATLAB, příp. naprogramujte ve vybraném programovacím jazyce.

Co se odevzdá

V referátu ve formátu PDF slovně komentujte vaše řešení, vč. zdůvodnění použitých předpokladů. Součástí referátu bude právě jedna tabulka a jeden graf dle zadání. Spolu s referátem odevzdejte pro posouzení komentovaný programový kód, který byl k řešení použit. Dbejte na splnění všech bodů zadání.