k-dimensionální strom (*kd*-tree)
(doplněk k IBL algoritmům)

Prohledávání prostoru mívá obecně vysokou výpočetní složitost (nelineárně narůstají požadavky na čas a paměť v závislosti na počtu zkoumaných bodů prostoru, n). Efektivnější hledání lze dosáhnout vymezením podprostoru. Jednou z možností je pomocí nadrovin (rovnoběžných s příslušnými osami) vymezit část k-rozměrného prostoru, která buď přímo obsahuje hledané řešení, nebo umožňuje již přijatelné libovolné dohledání (např. sekvenčně, pokud již prostor není rozdělen na jednoprvkové podprostory).

kd-strom je jedna z možností, kdy je prostor rozdělován postupným zpracováním jednotlivých os $x_1, x_2, \ldots, x_k, \ldots, x_1, x_2, \ldots, x_k$ tak, že se za ***bod dělení*** (jímž prochází příslušná nadrovina, kolmá k dané ose) volí ***medián*** souřadnic bodů v příslušném podintervalu vzniklém rozdělením předchozího intervalu. Zmíněný postup vede k vyváženému stromu, rozdělujícímu prostor až na podprostory obsahující 1 bod. U metody *nejbližšího souseda* l-NN pak stačí zjistit, ve kterém vzniklém podprostoru je zároveň l instancí známých (trénovacích) a ta neznámá (klasifikovaná). Navíc lze postupně dělit při klasifikaci jen ty podprostory, v nichž je klasifikovaná instance, jejíž souřadnice $\boldsymbol{x}$ jsou známy, ale není známo její hledané označení (které se zjistí až po nalezení požadovaného počtu l známých nejbližších "sousedů" ve výsledném podprostoru—zde je označován *počet sousedů* pomocí l, aby nedošlo k záměně s k, jež v tomto případě znamená *počet dimenzí prohledávaného prostoru*).

Horní výpočetní složitost vytvoření *kd*-stromu je $O(n \log_2 n)$ pro n zadaných bodů (prostor a vznikající podprostory jsou "půleny" mediánem, proto $\log_2$).

Existuje řada modifikací generování *kd*-stromu, kdy lze volit i jiný bod dělení než medián, přičemž pak nemusí být zaručena vyváženost stromu.