

Cvičení 8. *Zatímco meteorologické služby si platí síť stanic a provoz superpočítačů, každý by si mohl naprogramovat web, který by zprůměroval několik předpovědí a vytvořil z nich vlastní. Jaké by to mělo výhody (kromě minimální práce a nákladů) a nevýhody? Jak by se mohly meteorologické služby takové konkurenci bránit (bez utajení dat)? Jak by tato strategie dopadla u domácích úkolů z pravděpodobnosti a statistiky?*

Řešení. Zdánlivě by průměr měl být nepřesnější. Někdy to tak opravdu bude. Záleží na tom, co je zdrojem chyb. Zde to nemusí být nezávislé odchylky s přibližně normálním rozdělením. Zahrnou se i nekvalitní zdroje. Zdroje (a rovněž chyby) nemusí být nezávislé, takže nakonec mohou informace kolovat v kruhu a výsledek může být velmi podivný. (To je velký problém u bankovních produktů.)

Na druhé straně se tím stírají rozdíly; taková předpověď by nás často nevarovala před extrémů, které mohou být důležité. Např. se mohou zdroje shodnout v tom, že je vysoká pravděpodobnost lokální povodně, ale neshodnou se na tom, kde přesně. Naivním zprůměrováním se toto důležité varování ztratí. Taková předpověď by nám moc neřekla; nepřidali jsme žádnou informaci.

Meteorologické služby mohou argumentovat, že od nich se vlastně dovíme více, i když občas chybně.

U domácích úkolů by tato metoda vedla ke zcela špatným výsledkům; zprůměrováním několika dobrých a několika chybných odpovědí bychom se málokdy strefili do tolerované chyby. Místo výběrového průměru můžeme vzít např. výběrový medián, pak ale u meteorologických předpovědí můžeme dostat nesmyslné nespojitosti, např. srážkovou oblačnost, která se objeví „z ničeho“.