Document Length Normalization

- Penalize a long doc with a doc length normalizer
  - Long doc has a better chance to match any query
  - Need to avoid over-penalization
- A document is long because
  - it uses more words ➔ more penalization
  - it has more contents ➔ less penalization
- Pivoted length normalizer: average doc length as "pivot"
  - Normalizer = 1 if |d| =average doc length (avdl)