How Can We Combine Many Features? (Learning to Rank)

- General idea:
  - Given a query-doc pair (Q,D), define various kinds of features Xi(Q,D)
  - Examples of feature: the number of overlapping terms, BM25 score of Q and D, p(Q|D), PageRank of D, p(Q|Di), where Di may be anchor text or big font text, "does the URL contain '~'?"….
  - Hypothesize p(R=1|Q,D)=s(X1(Q,D),…,Xn(Q,D), λ) where λ is a set of parameters
  - Learn λ by fitting function s with training data, i.e., 3-tuples like (D, Q, 1) (D is relevant to Q) or (D,Q,0) (D is non-relevant to Q)