

Q2.1) O ganho de desempenho se deve ao mapeamento com a função de tangente hiperbólica, que permite a saturação da saída do neurônio e evita que poucas entradas bem alinhadas, porém de magnitude alta, induzam um resultado incorreto. Assim, ficam sendo mais influentes as entradas com muitos pontos coincidentes aos maiores valores dos classificadores, em detrimento daquelas que possuem poucos pontos de grande valor sendo multiplicadas pelos elementos do respectivo classificador.

A diminuição do tempo de execução se deve ao fato de que a matriz de entrada e a matriz  $W$  reduziram de 784 para 500 (sem considerar o offset) seu número de colunas e linhas, respectivamente. Com isso, a quantidade de operações diminui significativamente, pois a quantidade de elementos a serem operados nestas matrizes é grande e seu número foi reduzido em pouco mais de 1 terço.

Q2.2) Como a matriz  $H$  é produzida através da multiplicação entre a entrada  $X$  e a matriz  $V$  de números normalmente distribuídos, o posto de  $H$  é maior que o de  $X$ , o que faz com que a necessidade de regularização no caso da ELM seja menor. Como calculamos a regularização de forma a penalizar o crescimento dos vetores de  $W$ , o coeficiente de regularização será o menor possível para realizar a tarefa. Portanto, o coeficiente de regularização utilizado no classificador linear (675.588) foi maior do que para a ELM (64.000).

Q2.3) Como a matriz de entrada será diferente para cada caso dos pesos aleatórios, o coeficiente de regularização pode ser alterado, para mais ou para menos, dentro de uma faixa de valores que é relativa ao desvio padrão escolhido para a distribuição normal dos pesos aleatórios.