

Lista de Exercícios 2 – TP555 Inteligência Artificial e Machine Learning

Aluno: Bruno Ferreira Gomes

Matrícula:842

Ex1)

Já que é proposto uma solução cujo número de features é muito alto, a possível técnica seria a variante **SGD (Gradiente Descendente Estocástico)**. A técnica é capaz de localizar os pontos com menores erros calculando-se o gradiente de uma função. Com passos de cálculo denominados interações, e a partir de um ponto aleatório no domínio da função de erro, este algoritmo consegue chegar no ponto de mínimo da função, sendo esta onde o gradiente é nulo. Como padrão, a solução do erro quadrático médio remete a uma função cuja superfície representa um prato convexo, assim, a cada passo da interação, o valor do erro vai diminuindo até encontrar o ponto de mínimo global, e assim retornar os valores ótimos dos parâmetros. Este método é iterativo e faz a aproximação levando em conta o gradiente calculado com um único exemplo de treinamento.

Esta técnica é a escolhida devido ao fato de que a cada passo de interação, vários exemplos do treinamento não são mais necessários, assim chegando a um ponto mais específico, utilizando menos recursos computacionais. Devido a este fato, a convergência do algoritmo pode não acontecer quando o passo de aprendizagem é fixo, fazendo o erro ficar pequeno, porém sem estar necessariamente no ponto ótimo.

Ex2)

Por motivo da falta de padronização entre as amostras, os valores de erro quadrático médio seriam afetados pois estes utilizam de uma superfície que calcula a distância euclidiana entre os pontos. Com estes valores em escalas diferentes, este cálculo levaria a uma grande deformação da superfície, deste modo alterando os valores de erro, confundindo o ponto ótimo. Assim, todas as técnicas de regressão linear que utilizem este método, e são aplicadas amostras com falta de escalonamento, o algoritmo enfrentará problemas de convergência correta.

Para solução de tal problema, existem meios capazes de realizar o escalonamento destas amostras, deixando-as padronizadas em um intervalo específico, para assim, o processo do algoritmo ser aplicado e obter-se os valores corretos e esperados, existindo agora um ponto ótimo para convergir, pois a irregularidade na superfície de erro pôde ser corrigida.

Ex3)

Para solução de tal problema, deve-se atentar ao passo de aprendizado aplicado. Como o problema é o aumento do erro, deve-se escolher o valor certo para o passo de aprendizado sendo este não muito grande, para evitar o rápido aprendizado do algoritmo, fazendo com que este treinamento seja instável, e também não muito pequeno, para que o treinamento não seja demorado, fazendo com que o algoritmo nunca atinja o mínimo ou fique preso em outro ponto local.

Para solucionar estes problemas, deve-se utilizar métodos de ajuste do passo de aprendizagem como o ajuste com variação programada ou com variação adaptativa.

A variação programada é capaz de diminuir a o valor do passo de aprendizagem a cada época, durante o processo de treinamento.

A variação adaptativa é capaz de variar o passo de aprendizagem de acordo com a performance do modelo, e é capaz de selecionar e atualizar diferentes passos para cada parâmetro do modelo.

Ex4)

O algoritmo que seria capaz de alcançar a vizinhança do ponto ótimo mais rapidamente seria o **estocástico**, pois, ele utiliza apenas um exemplo para o treinamento para calcular o gradiente. Desta forma, tendo os atributos e rótulos obtidos sequencialmente, o processamento do algoritmo é muito mais rápido, chegando mais próximo ao ponto ótimo, mas podendo não obter a convergência para este.

Para garantir a convergência, deve ser utilizado o **batch**. Pois este calcula o gradiente de forma retilínea na projeção da superfície, garantindo que o ponto de mínimo será atingido. Este utiliza no processo todos os exemplos do conjunto de treinamento a cada iteração, tornando o processamento do algoritmo mais lento.

Para buscar a melhor convergência do estocástico e do mini-batch, devem ser aplicados ajustes no passo de aprendizagem, utilizando de variações programadas ou adaptativas para chegar em valores de α que garantam essa convergência. Para o mini-batch ainda, deve ser considerado o tamanho utilizado, tendo em vista que com o aumento do tamanho utilizado, a taxa de convergência é reduzida. Com uma implementação inteligente, o tamanho do mini-batch pode deixar de alterar a taxa de convergência.