

1.3 Margem máxima para problemas linearmente separáveis

Em problemas de classificação linearmente separáveis, a margem de separação de um separador linear é definida como a média das distâncias dos pontos mais próximos de cada classe ao separador. Assim, para que a margem de separação seja maximizada o separador deve estar no ponto médio entre os dois pontos. O separador linear de margem máxima será a mediatriz do segmento de reta que une os pontos mais próximos. Assim, para um problema de classificação binária com N_1 amostras para a classe C_1 e N_2 para a classe C_2 , há no plano N_1N_2 mediatrizes candidatas a serem um separador de margem máxima. Este número será menor quando são considerados somente aquelas mediatrizes que separam os dois conjuntos; entre elas, a de maior margem será a solução ótima segundo esta representação do problema.

Quando as mediatrizes são consideradas, suas equações podem ser obtidas diretamente dos pontos que as definem, sem necessidade de treinamento. Neste caso, ao invés de ajustar os pesos visando à minimização do erro e maximização da margem, o problema se resume a escolher o par de pontos de classes opostas, ou a mediatriz correspondente, que minimize o erro e maximize a margem de separação. Métodos como o Chipclas² consideram uma abordagem semelhante, porém, os pares de pontos que definem os separadores são determinados por um grafo cujas arestas, que definem as mediatrizes, existem somente entre amostras próximas no espaço, ou seja, somente pontos próximos da região de separação formam mediatrizes candidatas.

Neste problema, o aluno deverá:

1. Fazer uma revisão de literatura sobre o assunto.
2. Propor um método para selecionar o separador definido pelas mediatrizes.
3. Comparar a sua proposta com outros métodos, como o Perceptron e a Regressão Logística, considerando dados sintéticos bidimensionais.
4. Extrapolar o método proposto para o caso multi-dimensional e comparar os resultados com Perceptron e Regressão Logística, considerando pelo menos 5 conjuntos de dados reais e que sejam linearmente separáveis.

O formato do trabalho deve ser:

- **Formato:** Formato de artigo (Ref: padrão IEEE), duas colunas, mínimo de 5 e máximo de 8 páginas.
- **Estrutura:** O texto do trabalho deve conter pelo menos Título, Introdução, Revisão de literatura, Metodologia, Resultados, Discussões, Conclusões e Bibliografia.
- **Entrega:** via Moodle na data indicada.

²Torres, L. C. B., et al. "Distance-based large margin classifier suitable for integrated circuit implementation." *Electronics Letters* 51.24 (2015): 1967-1969.