Grupo: *Júlio Cesar Carreira Cury, Taymeson Furtado Chaves, Olívio do Nascimento Ferreira e Pedro Vieira Armond*

Relatório com respostas dos campos ‘To-Do’ do Notebook.

Professor: Eduardo Luz (DECOM)

Disciplina: Inteligência Computacional

**1 - Avaliando um modelo gigante**



**2 - Avaliação do modelo ‘Giant’ conforme modelos anteriores**

**ToDo: Avalie o seu modelo treinado conforme foi feito nos exemplos anteriores (10pt)**

O gráfico utiliza uma escala logarítmica para as épocas, permitindo que observemos o comportamento dos modelos tanto no início quanto em fases mais avançadas do treinamento. As curvas representam o erro de treinamento e validação para cinco modelos com diferentes tamanhos: Tiny, Small, Medium, Large e Giant. As linhas sólidas indicam o erro no treinamento, enquanto as linhas tracejadas mostram o erro na validação. O objetivo da análise é identificar o comportamento de cada modelo em termos de aprendizado, overfitting (quando o modelo se ajusta demais aos dados de treinamento) e underfitting (quando o modelo não consegue capturar a complexidade dos dados).

Começando pelo modelo Tiny (Tiny Train e Tiny Val), que está representado pela cor azul. A curva de erro de treinamento para o Tiny model desce de forma relativamente lenta, e ainda mantém um erro relativamente alto, mesmo após muitas épocas. O mesmo comportamento é observado na curva de validação, onde o erro não apresenta grandes melhorias. Isso indica que o modelo Tiny está sofrendo de underfitting, ou seja, ele não tem capacidade suficiente para capturar as nuances e complexidades dos dados, resultando em um ajuste ruim tanto no conjunto de treinamento quanto no conjunto de validação. Isso pode ser explicado pelo fato de que o modelo tem poucas camadas ou neurônios, o que limita sua expressividade e capacidade de generalização.

O Small Model, representado pela cor laranja, mostra uma melhoria mais rápida em relação ao Tiny model. Sua curva de erro de treinamento desce mais rapidamente, indicando que ele tem uma capacidade ligeiramente maior para capturar padrões nos dados. No entanto, o erro de validação também estabiliza em um nível relativamente alto, mostrando que, embora seja melhor que o Tiny model, o Small model ainda não tem capacidade suficiente para alcançar um desempenho significativamente melhor. A diferença entre os erros de treinamento e validação também é pequena, o que indica que, assim como o Tiny model, o Small model também está sofrendo de underfitting, embora em menor grau.

O Medium Model, representado pela cor verde, apresenta um comportamento bastante diferente. A curva de erro de treinamento cai rapidamente e atinge um erro muito baixo em poucas épocas. A curva de validação também desce de maneira constante e acompanha bem a curva de treinamento, o que indica que o modelo está aprendendo padrões dos dados sem sofrer de overfitting. Isso sugere que o Medium model tem uma capacidade de ajuste adequada, capturando a complexidade dos dados de forma eficaz, sem superajustar o modelo aos dados de treinamento. Este modelo parece ser o mais equilibrado entre capacidade e generalização, sendo capaz de performar bem tanto no conjunto de treinamento quanto no conjunto de validação.

O Large Model, representado pela cor vermelha, mostra uma queda ainda mais rápida no erro de treinamento, atingindo um valor muito baixo em poucas épocas. No entanto, a curva de validação, embora inicialmente siga a de treinamento, começa a estabilizar em um certo ponto, e a diferença entre os erros de treinamento e validação começa a aumentar. Isso indica que o Large model tem uma capacidade de ajuste muito alta, permitindo que ele se ajuste muito bem aos dados de treinamento. No entanto, a estabilização do erro de validação indica que ele começa a sofrer de overfitting, o que significa que o modelo está capturando não apenas os padrões dos dados, mas também o ruído específico do conjunto de treinamento. Como resultado, ele não generaliza tão bem para novos dados.

O Giant Model, representado pela cor roxa, segue um padrão semelhante ao Large model, mas de forma mais extrema. O erro de treinamento cai muito rapidamente, quase atingindo o valor mínimo possível, o que demonstra que o modelo tem uma capacidade de ajuste extremamente alta. No entanto, a curva de validação se estabiliza e, em algumas partes, parece piorar, indicando que o modelo está sofrendo de um overfitting significativo. O Giant model tem tantas camadas e neurônios que ele acaba memorizando os dados de treinamento, e sua capacidade de generalização para novos dados é muito limitada. Isso é típico de modelos com grande número de parâmetros, que, embora possam alcançar um erro de treinamento muito baixo, tendem a falhar na validação devido à sua tendência de superajustar os dados.

Em termos de comportamento geral, podemos ver que os modelos menores (Tiny e Small) estão sofrendo de underfitting, enquanto os modelos maiores (Large e Giant) sofrem de overfitting. O Medium model parece ser o mais balanceado, conseguindo aprender bem os padrões dos dados sem se ajustar excessivamente ao conjunto de treinamento. Para melhorar o desempenho dos modelos maiores, seria necessário aplicar técnicas de regularização, como Dropout ou L2 regularization, para reduzir o overfitting. Outra abordagem seria utilizar early stopping, onde o treinamento é interrompido quando o erro de validação começa a aumentar, evitando que o modelo memorize os dados de treinamento.

Em resumo, o gráfico mostra claramente como o tamanho do modelo afeta sua capacidade de ajuste e generalização. Modelos pequenos têm dificuldade em capturar padrões complexos nos dados, resultando em underfitting. Por outro lado, modelos muito grandes têm uma capacidade de ajuste excessiva, o que resulta em overfitting. O modelo médio parece estar em um ponto ideal, equilibrando bem a capacidade de aprendizado com a generalização. Essa análise reforça a importância de ajustar o tamanho do modelo e aplicar técnicas de regularização para obter um desempenho ideal, sem cair nos problemas de underfitting ou overfitting.

**3 - Avaliação por TensorBoard:**

### ****ToDo:**** Análises (10pt)

Quais análises você pode fazer sobre o TensorBoard?

Primeiro gráfico "epoch\_accuracy": esse gráfico apresenta a acurácia durante o treinamento e a validação para vários modelos. Observamos que o modelo gigante, representado como "Giant", tem uma acurácia de treinamento extremamente alta (quase 1.0) desde o início, o que indica que ele está se ajustando rapidamente aos dados de treinamento. No entanto, a acurácia de validação para esse modelo não apresenta uma melhora correspondente e se estabiliza em torno de 0.65, sugerindo um caso claro de overfitting. Isso significa que o modelo gigante está memorizando os dados de treinamento e não consegue generalizar bem para novos dados, o que é evidenciado pela diferença significativa entre a acurácia de treinamento e validação.

Os modelos menores, como o modelo "Tiny" e o modelo "Small", apresentam um comportamento oposto. Suas curvas de acurácia no treinamento são bem mais baixas, em torno de 0.60 para o modelo Small e 0.57 para o modelo Tiny, indicando que esses modelos têm uma capacidade limitada de aprendizado. As curvas de validação, embora ligeiramente melhores, também não indicam uma boa generalização, sugerindo que esses modelos sofrem de underfitting, onde eles não conseguem capturar adequadamente os padrões nos dados.

O modelo "Medium" apresenta uma curva de treinamento mais equilibrada. A acurácia de treinamento estabiliza em torno de 0.82, enquanto a acurácia de validação permanece em torno de 0.65, o que mostra que o modelo está aprendendo bem, embora ainda haja uma diferença entre treinamento e validação, o que sugere que ele pode estar começando a apresentar overfitting leve, mas ainda de forma mais controlada.

O segundo gráfico "epoch\_binary\_crossentropy" mostra a métrica de erro (binary crossentropy) ao longo do tempo para esses mesmos modelos. No caso do modelo Giant, o erro no conjunto de treinamento cai drasticamente para quase zero, indicando que ele está memorizando os dados de treinamento. Entretanto, no conjunto de validação, o erro permanece muito mais alto, em torno de 0.6, o que reforça a hipótese de overfitting. Esse comportamento é visto de forma similar no modelo Large, embora o erro de validação seja ainda maior, chegando a 1.57, o que indica que esse modelo também sofre de overfitting severo.

Os modelos menores, como o Small e o Tiny, mostram um erro de treinamento e validação que são próximos, mas relativamente elevados, por volta de 0.55 e 0.57, respectivamente. Isso indica que esses modelos têm uma capacidade limitada de aprendizado e sofrem de underfitting, ou seja, eles não conseguem capturar padrões suficientes dos dados de treinamento para fazer previsões precisas.

O modelo "Medium" apresenta uma queda mais gradual no erro de treinamento, que converge para um valor razoável de 0.3, enquanto o erro de validação permanece por volta de 0.48. Isso sugere que o modelo está aprendendo com os dados de treinamento, mas ainda comete erros consideráveis no conjunto de validação, o que pode ser uma indicação de overfitting moderado.

O gráfico "epoch\_learning\_rate" mostra a taxa de aprendizado decrescendo ao longo do tempo. Todos os modelos parecem seguir um padrão semelhante de redução da taxa de aprendizado, começando com uma taxa mais alta e diminuindo à medida que o treinamento avança. Isso indica que a estratégia de ajuste da taxa de aprendizado é consistente para todos os modelos, o que ajuda na convergência do treinamento. Entretanto, para os modelos Giant e Large, a redução da taxa de aprendizado não foi suficiente para evitar o overfitting, indicando que outras estratégias de regularização, como o uso de dropout ou early stopping, poderiam ser mais eficazes.

No gráfico "epoch\_loss", vemos novamente a métrica de erro ao longo do treinamento. O modelo Giant mostra uma perda muito baixa no conjunto de treinamento, caindo quase para zero, o que reforça a ideia de overfitting. O erro de validação, no entanto, permanece muito mais alto, por volta de 2.5, o que confirma a incapacidade desse modelo de generalizar para novos dados. Um comportamento semelhante é observado no modelo Large, onde a perda de treinamento é extremamente baixa e a perda de validação é elevada, acima de 2.3.

Por outro lado, os modelos menores, como Small e Tiny, apresentam erros tanto no treinamento quanto na validação que são muito mais altos e próximos entre si. Isso sugere que esses modelos têm uma capacidade limitada de ajuste e não conseguem capturar padrões adequados dos dados, levando a um subajuste, ou underfitting.

O gráfico "evaluation\_accuracy\_vs\_iterations" apresenta a acurácia de avaliação ao longo das iterações. Observa-se que o modelo Medium mantém uma acurácia de validação em torno de 0.65, sugerindo que ele está bem ajustado. No entanto, o modelo Giant e Large, que apresentam flutuações significativas na acurácia de validação, reforçam a noção de overfitting, já que não conseguem manter uma acurácia estável no conjunto de validação. Os modelos Small e Tiny também não apresentam uma acurácia alta, indicando underfitting.

Finalmente, no gráfico "evaluation\_binary\_crossentropy\_vs\_iterations", o erro de validação é novamente destacado. Os modelos Giant e Large apresentam erros elevados, confirmando overfitting. O modelo Medium, embora com uma diferença menor entre treinamento e validação, ainda tem um erro de validação relativamente alto, o que indica que ele poderia se beneficiar de ajustes adicionais. Já os modelos Small e Tiny, com erros altos e próximos, continuam a mostrar sinais claros de underfitting.

Conclusão geral: os gráficos revelam que os modelos Giant e Large sofrem de overfitting severo, enquanto os modelos Small e Tiny sofrem de underfitting. O modelo Medium parece ser o mais equilibrado entre eles, mas ainda apresenta sinais de overfitting moderado. Estratégias de regularização mais fortes, como dropout, penalização L2, ou o uso de early stopping, poderiam ser aplicadas aos modelos maiores para evitar o overfitting. Além disso, aumentar a capacidade dos modelos menores poderia ajudar a resolver o problema de underfitting.

**4 - Análise da regularização:**

#### **ToDo:** Análise dos resultados (10pt)

O que você pode inferir analisando os resultados apresentados no TensorBoard?

O gráfico apresenta resultados de diferentes estratégias de regularização no treinamento de redes neurais. A análise pode ser feita com base nas métricas fornecidas: acurácia, entropia binária cruzada, taxa de aprendizado e perda.

Epoch Accuracy (acurácia por época): As curvas mostram que o modelo "Tiny" começa com uma taxa de acurácia menor em relação aos outros, mas rapidamente se estabiliza, com desempenho melhor na validação comparado aos demais. O modelo "combined" (L2 + Dropout) tem uma curva de aprendizado mais suave e estável, com uma melhora progressiva ao longo do tempo. O modelo "dropout" também apresenta consistência, enquanto o modelo "L2" evolui de maneira semelhante ao "dropout", mas com flutuações menores. O "Tiny" tem bons resultados, mas os ganhos são limitados comparado aos modelos regularizados. A combinação de regularização L2 e Dropout parece ser a estratégia mais eficiente para melhorar a acurácia e prevenir o overfitting, como demonstrado pela estabilidade e crescimento constante da acurácia ao longo do tempo. A regularização individual (L2 ou Dropout) também mostra impacto positivo, embora de forma menos expressiva.

Epoch Binary Crossentropy (entropia cruzada binária por época): A entropia binária começa alta para todos os modelos, mas os modelos "dropout" e "combined" apresentam quedas mais significativas, demonstrando melhor otimização da função de perda. O "Tiny" se estabiliza, mas mantém uma entropia cruzada maior em comparação aos modelos regularizados, especialmente na validação. O modelo "L2" também tem uma melhora significativa, com uma trajetória suave de redução da entropia. A combinação de regularização L2 e Dropout reduz drasticamente a entropia cruzada binária, mostrando que os modelos regularizados são mais eficazes na minimização da perda em relação ao "Tiny". Isso indica que esses modelos são menos propensos a overfitting, aprendendo representações mais generalizáveis.

Epoch Learning Rate (taxa de aprendizado por época): A taxa de aprendizado decresce gradativamente ao longo do treinamento para todos os modelos, comportamento típico ao usar otimizadores como Adam ou SGD com decay. Os modelos com regularização L2 e Dropout mantêm uma taxa de aprendizado controlada, permitindo ajustes finos nos pesos sem causar grandes variações. Uma taxa de aprendizado controlada garante que o modelo faça ajustes pequenos e constantes nas últimas etapas do treinamento, evitando overshooting e garantindo convergência estável, especialmente nos modelos com regularização.

Epoch Loss (perda por época): As curvas de perda seguem padrões semelhantes às da entropia binária, com os modelos "dropout" e "combined" apresentando uma queda rápida e estabilização em níveis mais baixos. O modelo "Tiny" apresenta flutuações mais pronunciadas, especialmente na validação, o que indica uma menor capacidade de generalização. O modelo "L2" também mostra uma queda acentuada na perda, embora seja mais estável que o "Tiny". A perda mais baixa nos modelos regularizados demonstra melhor ajuste aos dados de treinamento sem superajustar. O "Tiny" não consegue reduzir a perda de forma tão eficiente.

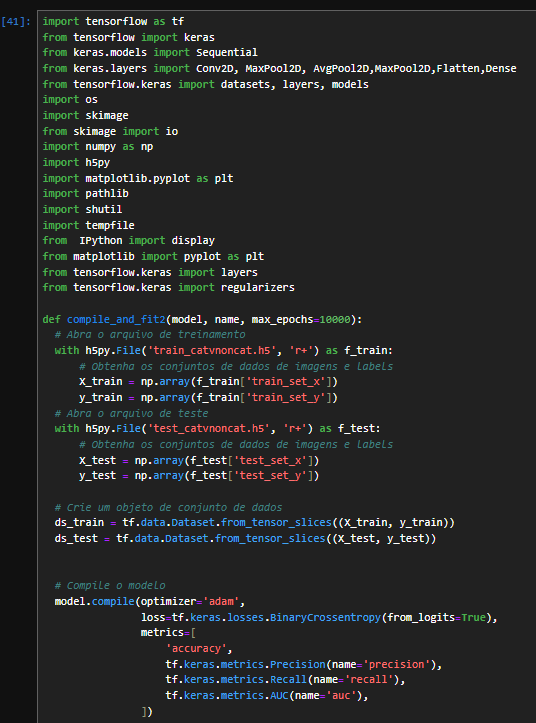
Evaluation Accuracy vs Iterations (acurácia de avaliação vs iterações): Conforme o número de iterações aumenta, o modelo "combined" mantém uma taxa de acurácia mais alta, sugerindo que ele se beneficia do aumento no número de iterações. O "Tiny" atinge uma acurácia consistente após certo número de iterações, mas não supera os modelos regularizados. O modelo "combined" lida melhor com um maior número de iterações, mantendo uma performance crescente, enquanto os demais modelos se estabilizam em níveis menores de acurácia.

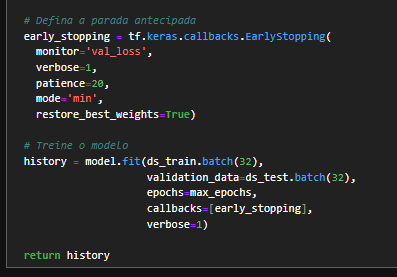
Evaluation Binary Crossentropy vs Iterations (entropia binária cruzada vs iterações): O modelo "Tiny" apresenta entropia binária mais alta ao longo das iterações, com maior flutuação. Os modelos regularizados, especialmente o "combined", apresentam uma redução significativa e estável na entropia cruzada. O modelo "combined" demonstra ser mais eficaz na redução da entropia ao longo das iterações, consolidando sua capacidade de generalização superior ao "Tiny".

Evaluation Loss vs Iterations (perda de avaliação vs iterações): O modelo "combined" tem a menor perda em comparação aos outros, enquanto o "Tiny" apresenta maior perda e flutuações durante as iterações. Os modelos "dropout" e "L2" conseguem manter uma perda baixa, mas o "combined" se destaca. A combinação de L2 e Dropout é a técnica mais robusta, mantendo a perda consistente e baixa ao longo do treinamento.

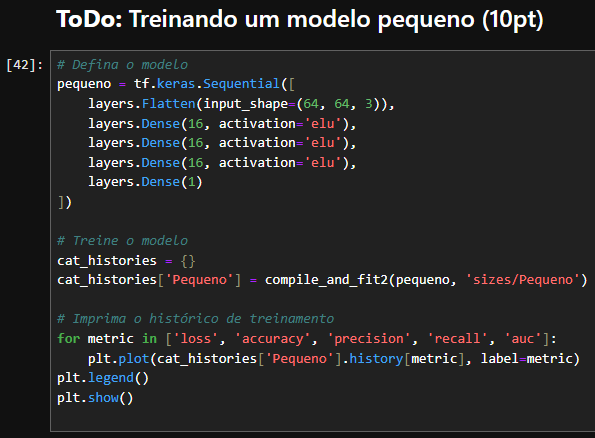
Conclusão geral: As técnicas de regularização, especialmente a combinação de L2 e Dropout, foram eficazes em reduzir o overfitting e melhorar o desempenho geral do modelo. O "Tiny", usado como base, tem boa performance inicial, mas rapidamente é superado pelos modelos regularizados, que apresentam melhores resultados em acurácia, perda e entropia cruzada, além de maior estabilidade durante o treinamento e avaliação.

**5 – Lendo o Dataset ‘Gatos-Não Gatos’:**





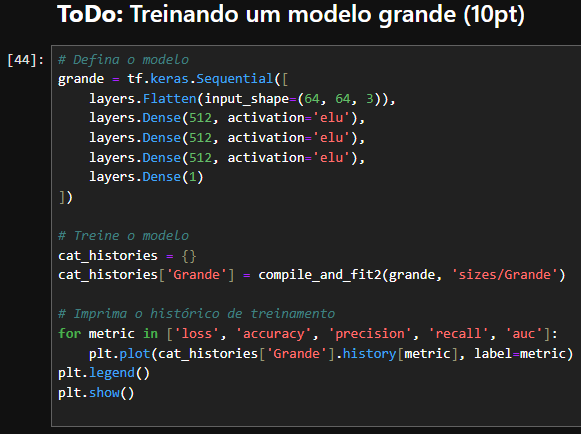
**6 – Treinamento modelo pequeno:**



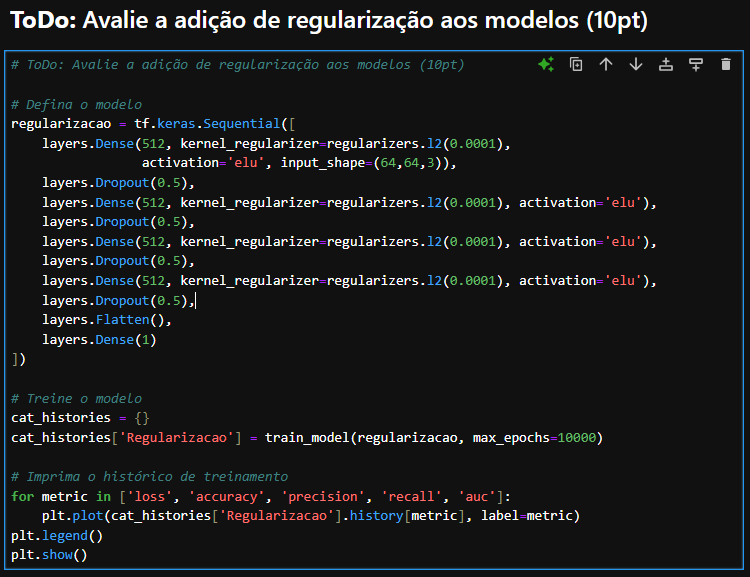
**7 – Treinamento modelo médio:**



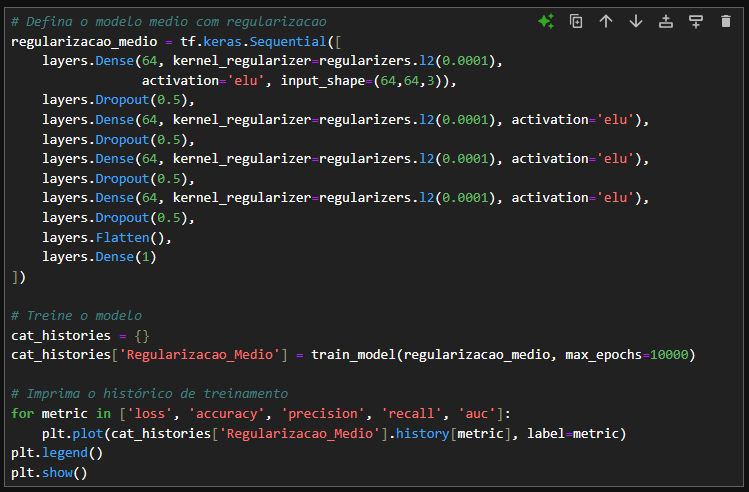
**8 – Treinamento modelo grande:**



**9 – Avaliação de Regularização no Modelo Gatos-Não gatos:**



**10 – Modelo médio com Regularização:**

****

**11 – Análise final dos resultados:**

