

# Avances 14 Julio

Juan David Medina Tobón

Universidad de los Andes

*jd.medinat@uniandes.edu.co*

14 de julio de 2020

# Problemas Actuales

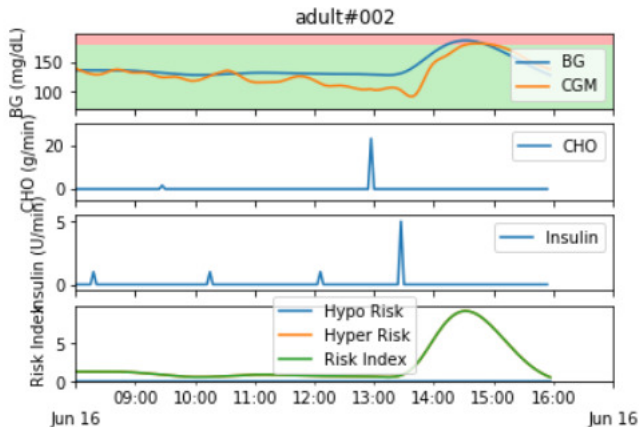


Figura: Escenario de Prueba

# Problemas Actuales

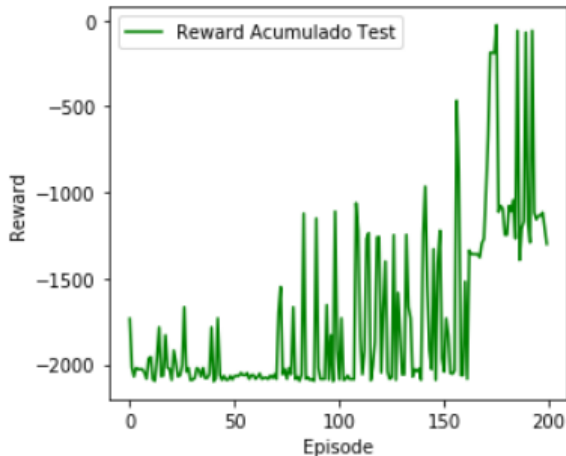


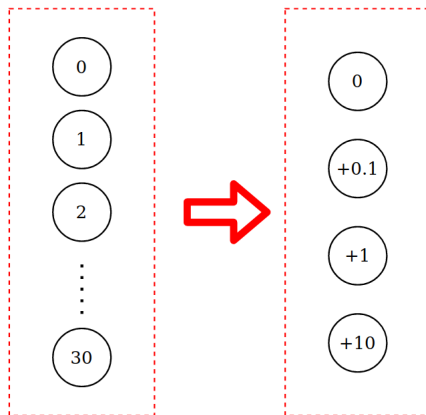
Figura: Recompensa Acumulada

El método de inicialización de los pesos de las redes se cambió por el recomendado en [1] para redes con activación ReLu.

$$w_l \sim N(0, \frac{1}{n_l}) \rightarrow w_l \sim N(0, \frac{2}{n_l})$$

# Cambios

Las acciones se plantearon inicialmente como una discretización en pasos de 1. Esto se cambió por desviaciones de diferente tamaño de la cantidad de insulina inyectada actual, por lo cual dicha cantidad actual se agregó al estado.



# Resultados

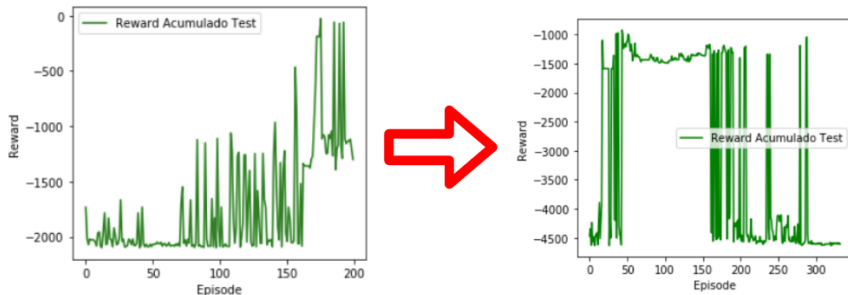


Figura: Resultados Recompensa Acumulada

# Resultados

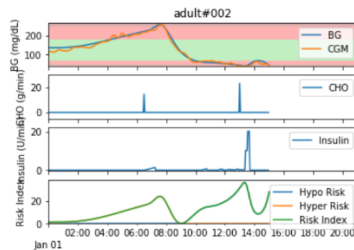
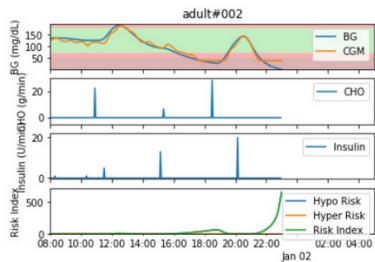


Figura: Resultados Escenario de Prueba

- Implementar técnicas mas avanzadas de Experience Replay como aquella propuesta en [2].
- Implementar métodos de exploración más adecuados para el problema como el algoritmo de Soft Q-Learning de [3].



- ① He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034)
- ② Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). Prioritized experience replay. arXiv preprint arXiv:1511.05952.
- ③ Haarnoja, T., Tang, H., Abbeel, P., & Levine, S. (2017). Reinforcement learning with deep energy-based policies. arXiv preprint arXiv:1702.08165.