

## POS tagging

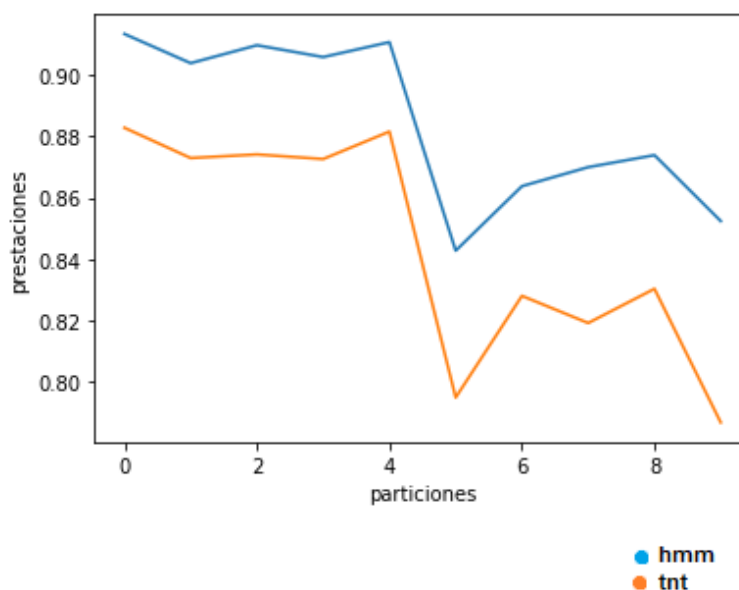
Aquí se mostrará los resultados del uso de NTK y Python para el procesamiento del corpus y obtener el POS tagging.

Primero descargamos el corpus NTK, reducimos el número de etiquetas de 289 a 66 y separamos los datos en 90% para realizar el entrenamiento y 10% para la evaluación. Después importamos dos tipos de modelos hmm y tnt. Obtenemos así un porcentaje de acierto de 85% para el modelo hmm y 79% para el modelo tnt.

Tras esto dividimos el corpus en 10 partes, asignando para la partición 1 la primera para evaluación y el resto para training, para la 2 la segunda para evaluación y el resto para training y así sucesivamente. Obteniendo:

Partición:	1	2	3	4	5	6	7	8	9	10
Hmm	91%	90%	91%	91%	91%	84%	86%	87%	87%	85%
Tnt	88%	87%	87%	87%	88%	79%	83%	82%	83%	79%

Que gráficamente queda representado como:

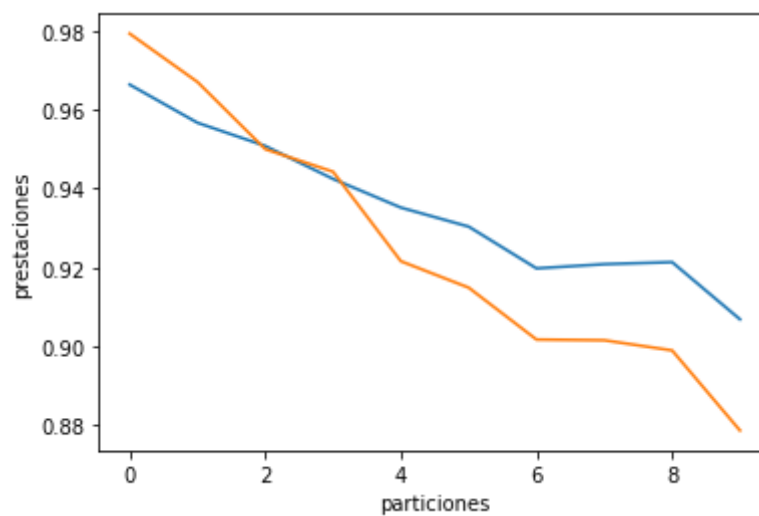


Vemos que obtenemos valores muy similares, si tuviéramos que quedarnos con alguna partición sería la 1 o la 5.

Después barajamos los datos y realizamos la evaluación de nuevo de la misma forma.  
Obteniendo:

Partición:	1	2	3	4	5	6	7	8	9	10
Hmm	97%	96%	95%	94%	93%	93%	92%	92%	92%	90%
Tnt	98%	97%	95%	94%	92%	91%	90%	90%	90%	88%

Que gráficamente queda:



Parece que al barajar los datos la partición 1 rinde mejor que las demás.

Gracias por su atención