**Evaluaciones**

En este capítulo se detallaran las diferentes evaluaciones realizadas sobre los datos para seleccionar la estrategia optima que maximice la correlacion entre la informacion histórica captada kilómetros adentro en el oceano y las observaciones visuales en la costa y que al mismo tiempo minimice los margenes de error. Para realizar estas evaluaciones utilizamos diferentes algoritmos de regresión (redes neuronales, maquinas de soporte vectorial, regresion lineal, procesos gaussianos) buscando el que mejores resultados generaba.

Basicamente, la información que utilizamos para entrenar a los clasificadores consiste en dos grandes grupos de datos. El primero es un historico de datos del modelo WW3, situados en cuatro coordenadas alrededor de OAHU a unos cuantos kilometros de la costa, del cual disponemos de lecturas del comportamiento del oceano (periodo de ola, altura de ola, dirección de ola, dirección del viento y velocidad del viento) cada 3 horas durante los 365 dias del año desde 1997 hasta la fecha. El segundo grupo es una base de datos de observaciones visuales de la altura de las olas de diferentes playas alrededor de toda la costa de Oahu desde el año 1987 hasta el año 2004 inclusive. Estas observaciones realizadas por un experto en el area nos brinda por cada dia del año la altura de la ola más grande que arribo a la costa ese dia. Además estas observaciones sabemos (ya que la fuente lo especifica) que fueron realizadas durante el periodo de luz solar de dia en esa playa.

A partir de estos dos grupos de datos, comenzamos a implementar diferentes estrategias de filtrado y utilizacion de los mismos hasta encontrar una que optimice los aciertos en la prediccion de las olas en las diferentes costas de la isla.

**Filtrado basico de la informacion**

Dado que las lecturas de WW3 se actualizan cada 3 horas, tanto durante el dia como la noche, y debido a que las observaciones visuales se han realizado en horarios en que se cuenta con luz solar, el primer filtro aplicado a la informacion que disponemos del modelo WW3 fue eliminar todas las lecturas que no correspondian con un horario de luz natural en Oahu.

Por otra parte, las observaciones realizadas nos indican la altura de la ola más grande que arribo a la costa cada dia del año, es decir que disponemos de una lectura diaria, por esto el segundo filtro aplicado a las lecturas de WW3 fue quedarse unicamente con la lectura diaria con la mayor altura de ola. De esta forma ambos conjuntos de datos cuentan con una lectura diaria que indica la ola más grande captada y durante una hora en Oahu en la que hay luz solar.

Los filtros anteriormente descriptos fueron los que finalmente se utilizaron ya que luego veremos en las evaluaciones que con estos obtuvimos los mejores resultados. Otros filtros que tuvimos en cuenta para realizar pruebas (descartados por no mejorar los resultados) fueron entre otros, utilizar las lecturas diarias del WW3 en el que la altura de la ola se acerca más al promedio de altura de olas de cada dia, filtrar las lecturas cuya dirección de ola no estaba dirigida a la costa a evaluar, agrupar las lecturas en dos grupos (dos rangos de meses) que distinguen la temporada de olas grandes de la temporada de olas pequeñas, etc.

**Evaluaciones**

Para encontrar a una estrategia óptima sobre los datos que maximice la correlación entre las lecturas del WW3 y las observaciones visuales, y a su vez minimicen los márgenes de error entre una futura predicción de altura de la ola y la altura real que esta alcanzara en la costa se aplicaron diferentes algoritmos y pre-procesamientos sobre los datos que vamos a ir presentando a continuación, en orden desde la peor estrategia hasta la estrategia escogida como la optima para entrenar al clasificador del sistema final. Siempre teniendo en cuenta que todas las estrategias a continuación utilizan como conjunto de entrada datos ya filtrados como se explicó en la sección anterior (Filtrado básico de la información).