

Instituto Tecnológico Superior de Misantla

Ingeniería En Sistemas Computacionales

Octavo semestre.

Arquitectura reconfigurable

**Unidad III**

Clasificación de imágenes.

**Alumno**

*Cornelio Ramirez Campos*

Docente: **Ing. Pablo Colorado Posadas.**

Fecha de entrega: **29 de julio del 2017**

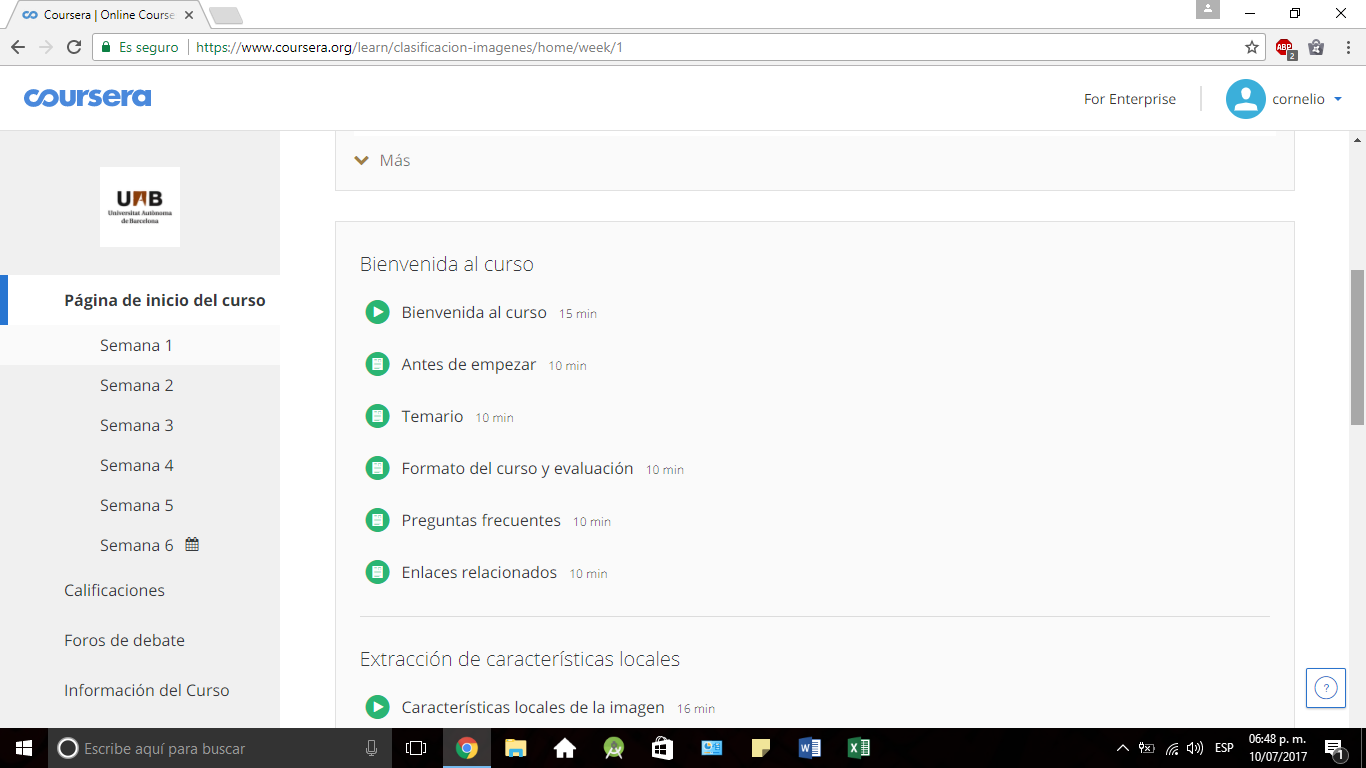
Clasificación de imágenes: ¿cómo reconocer el contenido de una imagen?

En la semana uno de este curso se explicó del contenido del curso y de los temas que se iban a enseñar en cada semana y como es que se iba a trabajar en el curso.

Se dio el concepto de imagen el cual nos dice que:

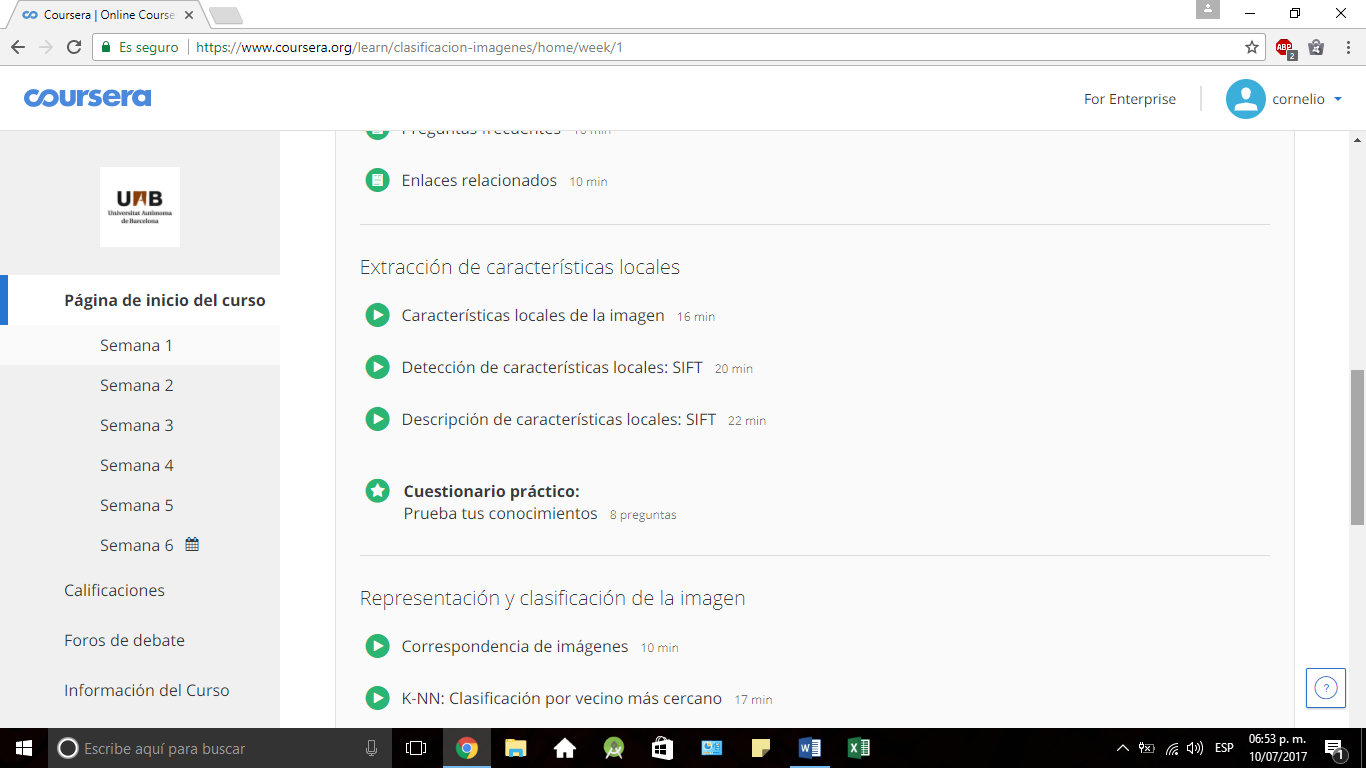
Es el resultado de la combinación de tres factores que intervienen en el proceso de captura de la imagen: El color de la luz, el material de la superficie de los objetos y la sensibilidad de la cámara.

Y lo que se vio es cómo podemos diseñar sistemas de clasificación de imágenes que sean capaces de aprender qué etiqueta tenemos que asignar a una determinada imagen, a partir de un conjunto de imágenes de aprendizaje que ya han sido previamente etiquetadas. Y cómo se extraen las características visuales en una imagen, cómo se representan, y cómo se aprende lo que es característico de cada categoría, es lo que se vio a lo largo del curso.

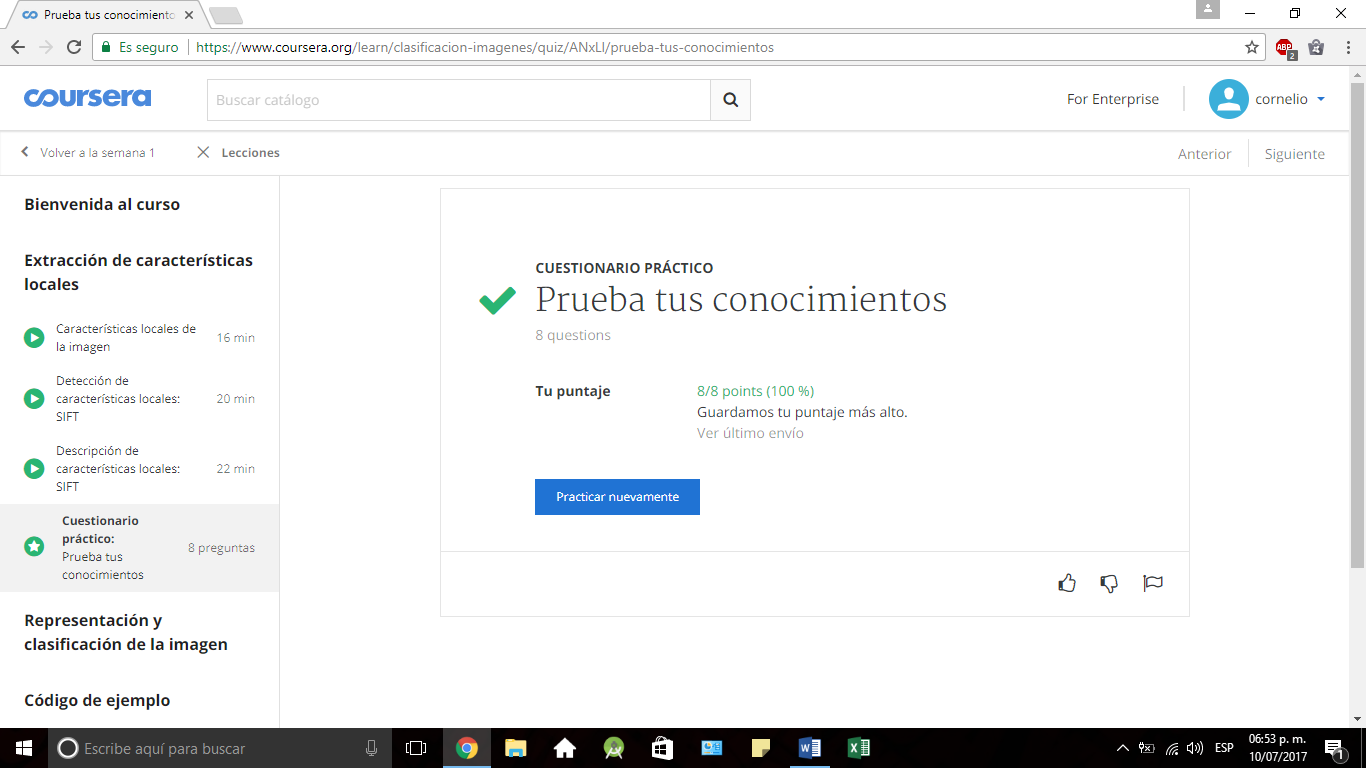


Una vez que se dio a conocer el formato de trabajo, se abordó el primer tema que fue el de extracción de características locales el cual incluían tres videos y un cuestionario práctico.

Para este método en primer lugar se detectan en la imagen los puntos de interés, es decir, zonas que puedan dar una información relevante sobre el contenido de la imagen.  Normalmente, estos puntos estarán localizados en zonas significativas. En segundo lugar, para cada uno de estos puntos de interés, se obtiene una descripción local. Es decir, una descripción del contenido visual en una pequeña zona de la imagen que definiremos alrededor del punto de interés. Esta descripción va a ser un vector numérico que, va a dar una representación matemática del contenido visual de esa zona.



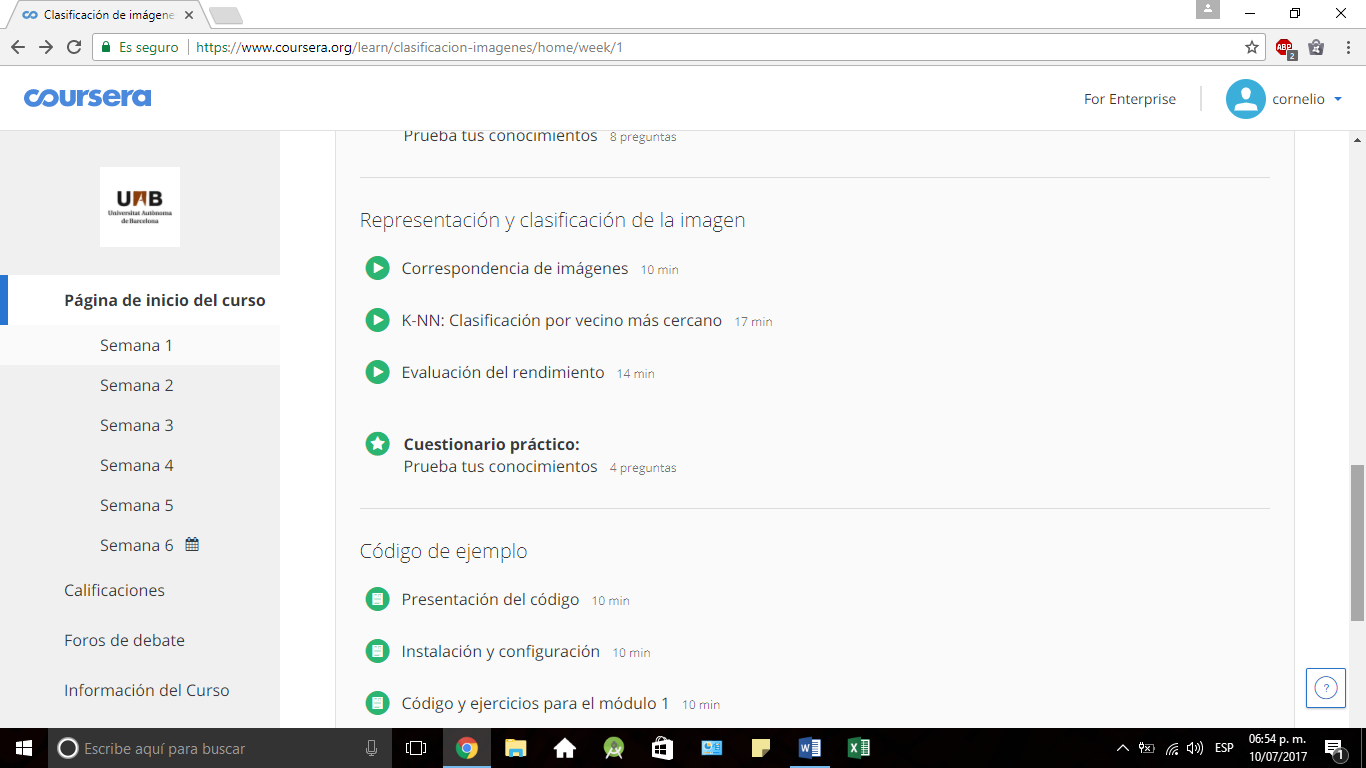
En el primer cuestionario que consto de 8 preguntas las cuales eran de los temas como convolución, filtros, gradiente y así como el descriptor SIFT.

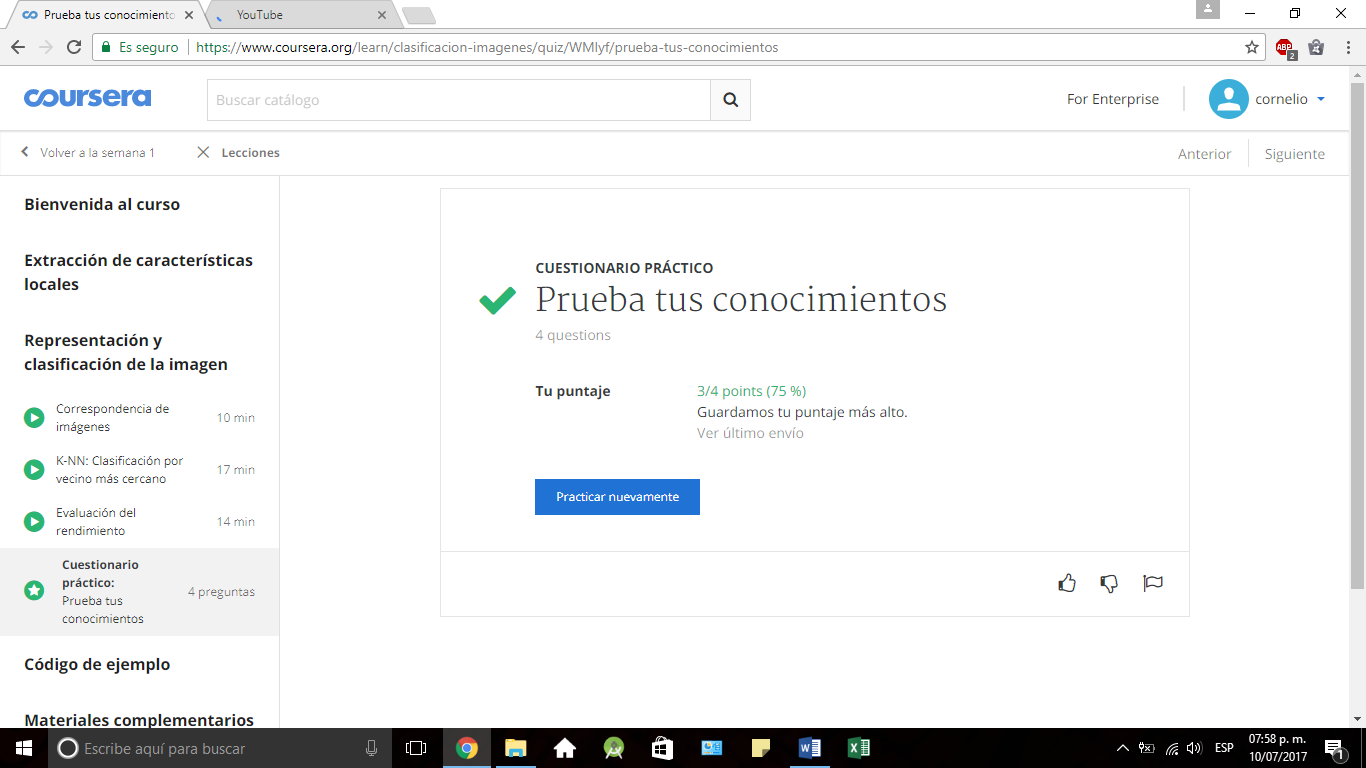


En el siguiente tema de la semana uno: Representación y clasificación de la imagen se habló de la clasificación por vecino más cercano, este tema consto de tres videos y un cuestionario practico.

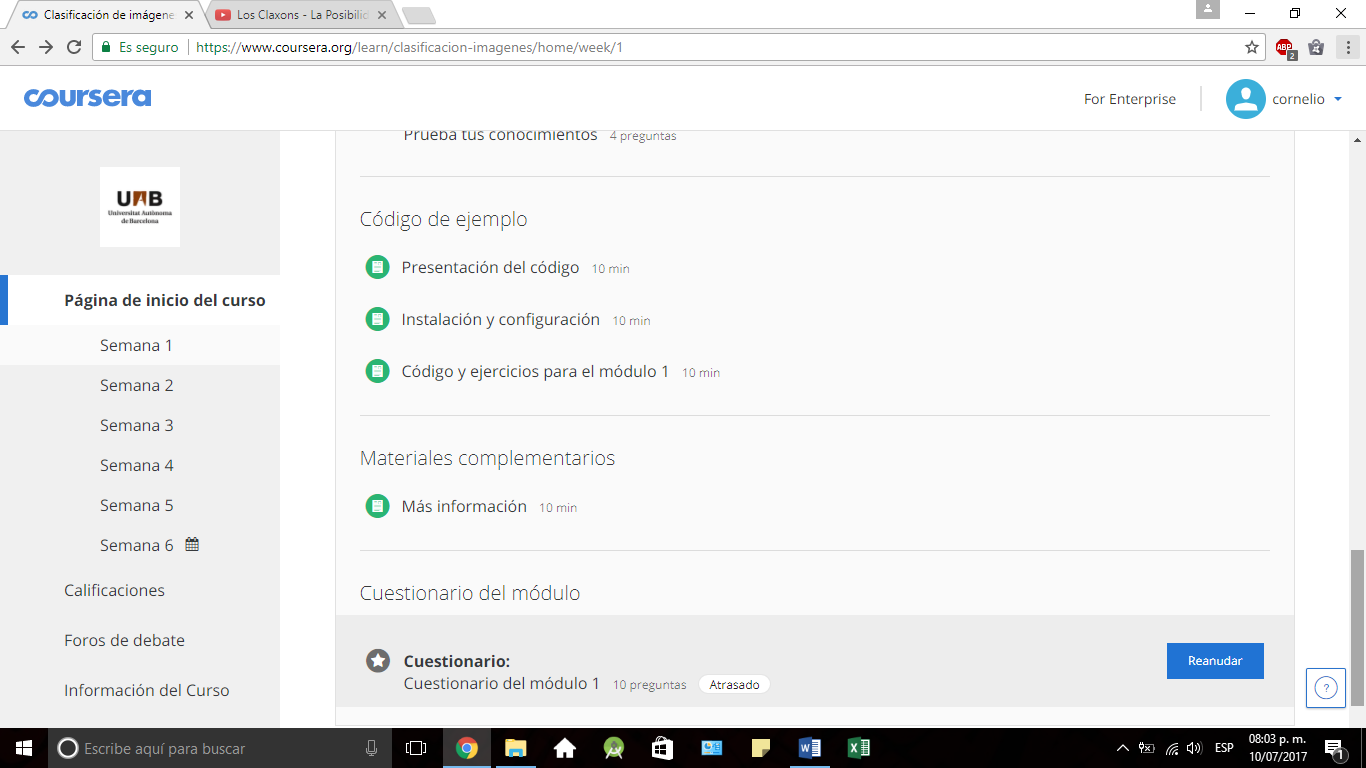
Después de la extracción de características el siguiente paso en el esquema será obtener una representación única de toda la imagen, en forma de vector numérico. Esta representación se obtiene agregando o combinando la descripción de todos los puntos de interés.

A partir de esta representación única de toda la imagen, el paso final va a ser el paso de clasificación. Es en este último paso donde se aprende, a partir del conjunto de aprendizaje, cómo distinguir las imágenes de las diferentes categorías a partir de la representación visual de cada imagen.

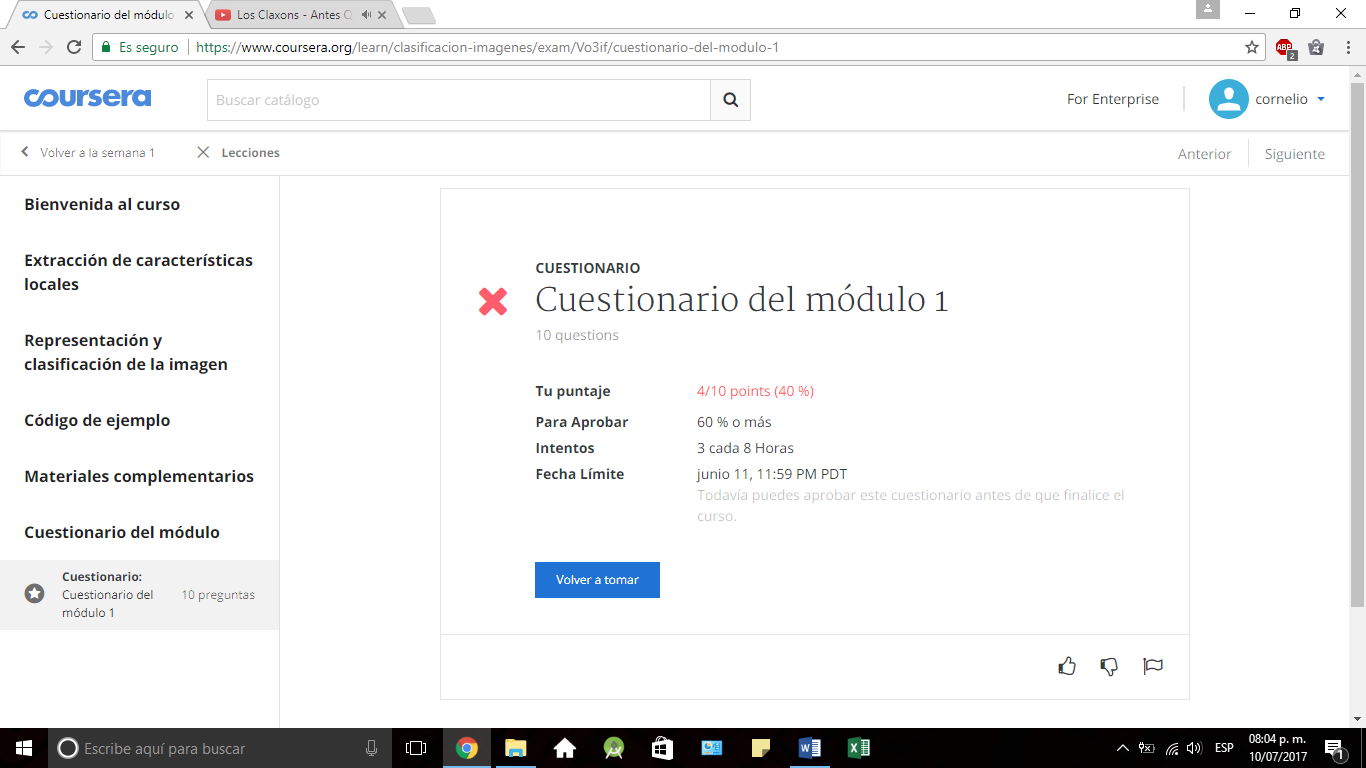


Una vez terminados los videos en el cuestionario practico se hicieron preguntas sobre el conjunto de entrenamiento y de muestras el cual solo consto de 4 preguntas.

Se dio un código ejemplo y se llevaron a cabo los pasos para instalar y configurar Python para la realización de unos ejemplos y ejercicios en Python, estos no influyen en las calificaciones del curso, es opcional y como una ayuda para comprender mejor del tema.



Ya una vez terminados los temas se realizó el cuestionario final de la primera semana el cual tenía 10 preguntas.

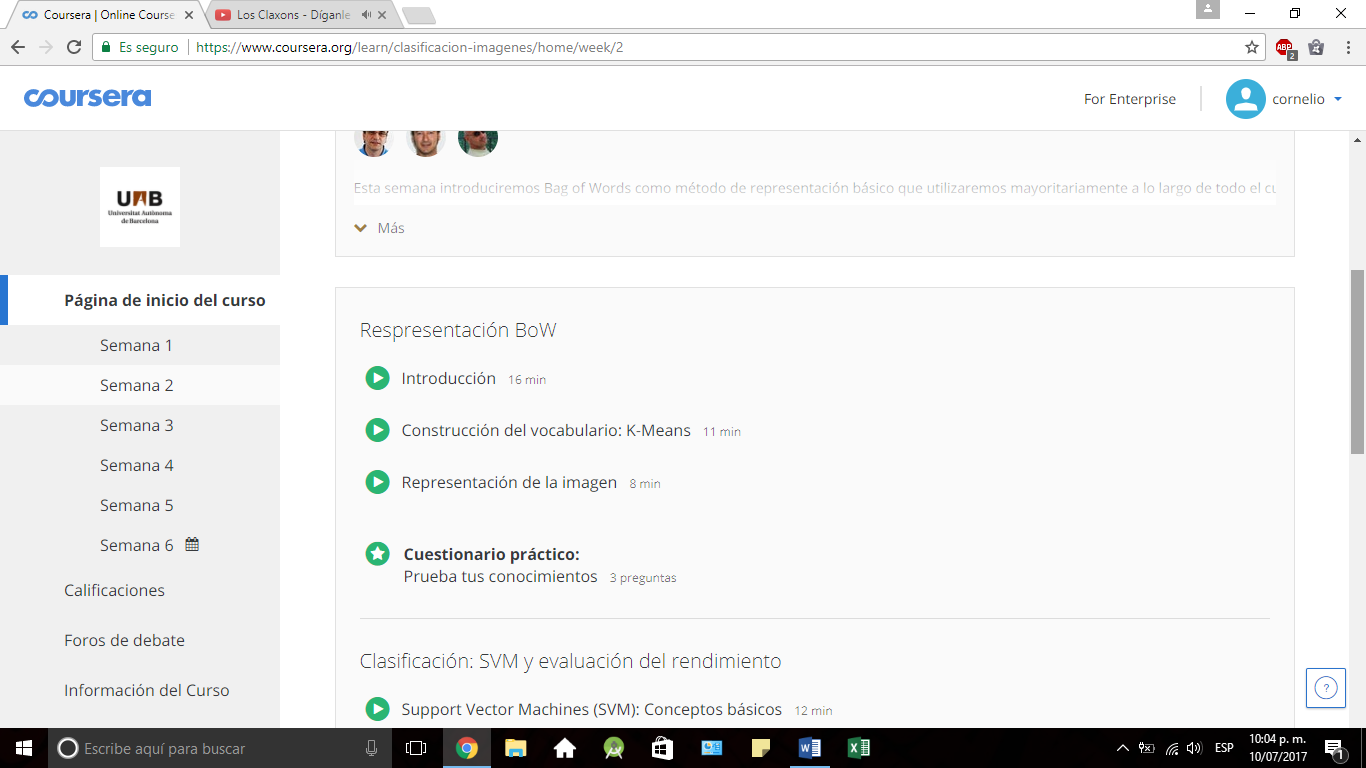


**SEMANA 2**

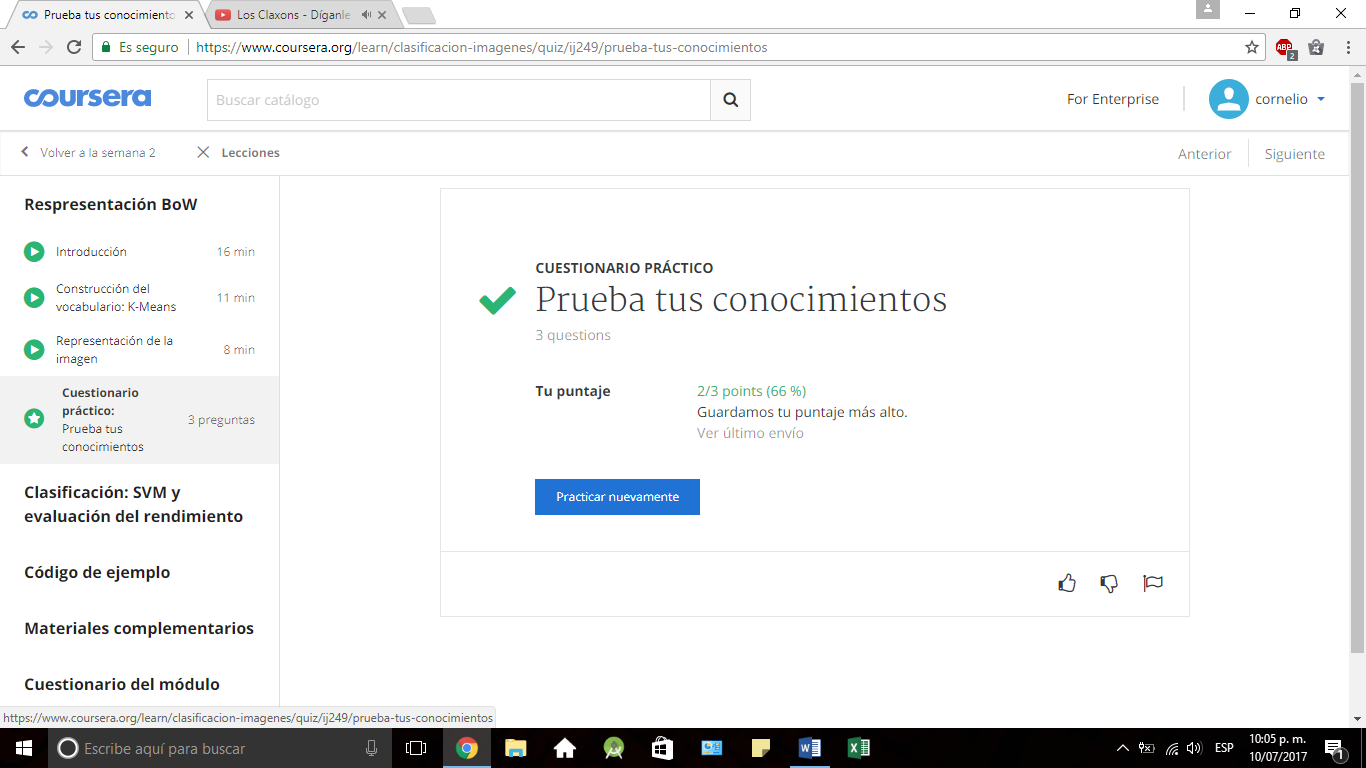
La segunda semana que consto de los temas de representación BoW y clasificación con SVM y evaluación del rendimiento.

El primer tema consto de tres videos y un cuestionario, en el primer video se dio la introducción del tema, en el segundo video hablo de la construcción del vocabulario K-means, el cual permite  encontrar grupos de características similares y poder determinar un representante para cada uno de los grupos que van a ser cada una de nuestras palabras visuales.

En el siguiente video se habló de la representación de la imagen siguiendo el esquema de clasificación que se ha ido viendo a lo largo del curso, esta representación se va a obtener a partir de las características locales de la imagen calculadas por ejemplo utilizando SIFT y las palabras visuales nos van a servir para discretizar o agrupar estas características de la imagen y así poder obtener un histograma de palabras como representación final de toda la imagen.

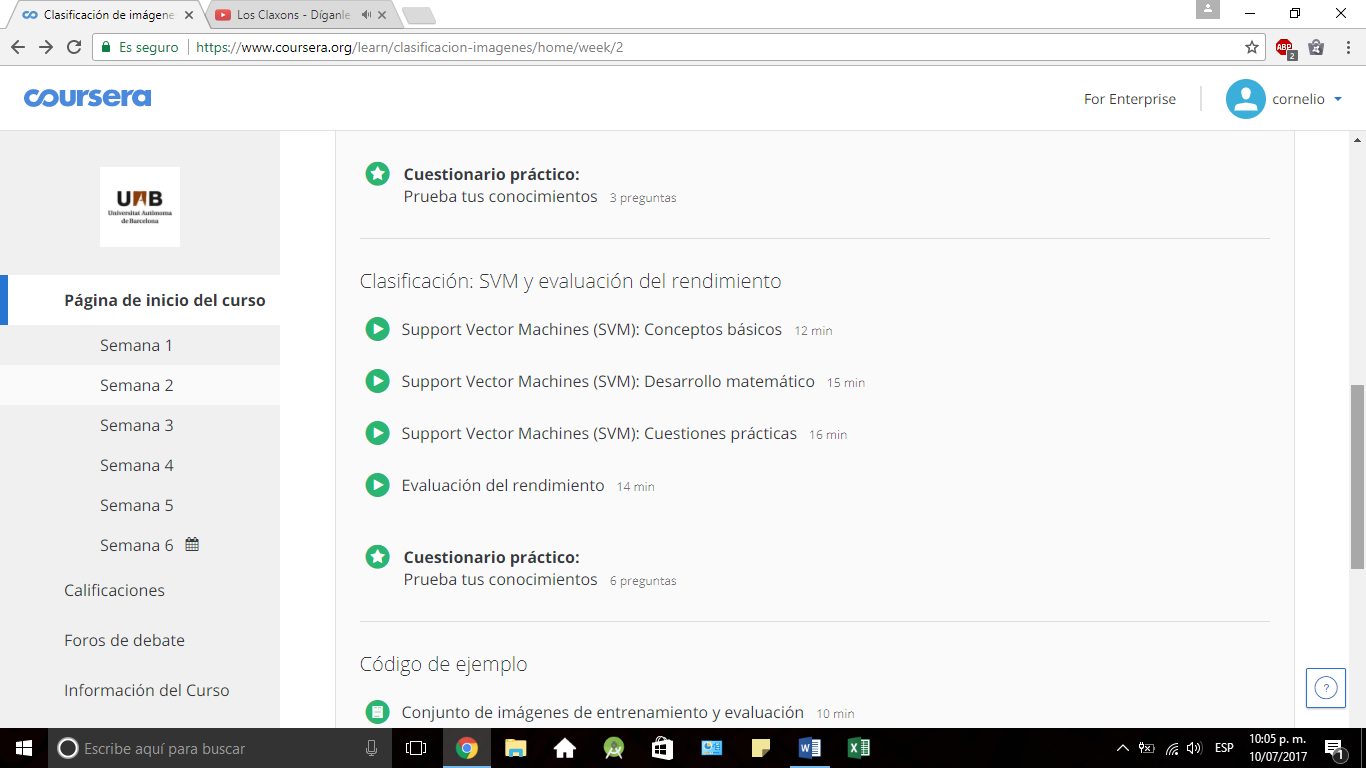


Se realizó el cuestionario práctico el cual constaba de 4 preguntas.

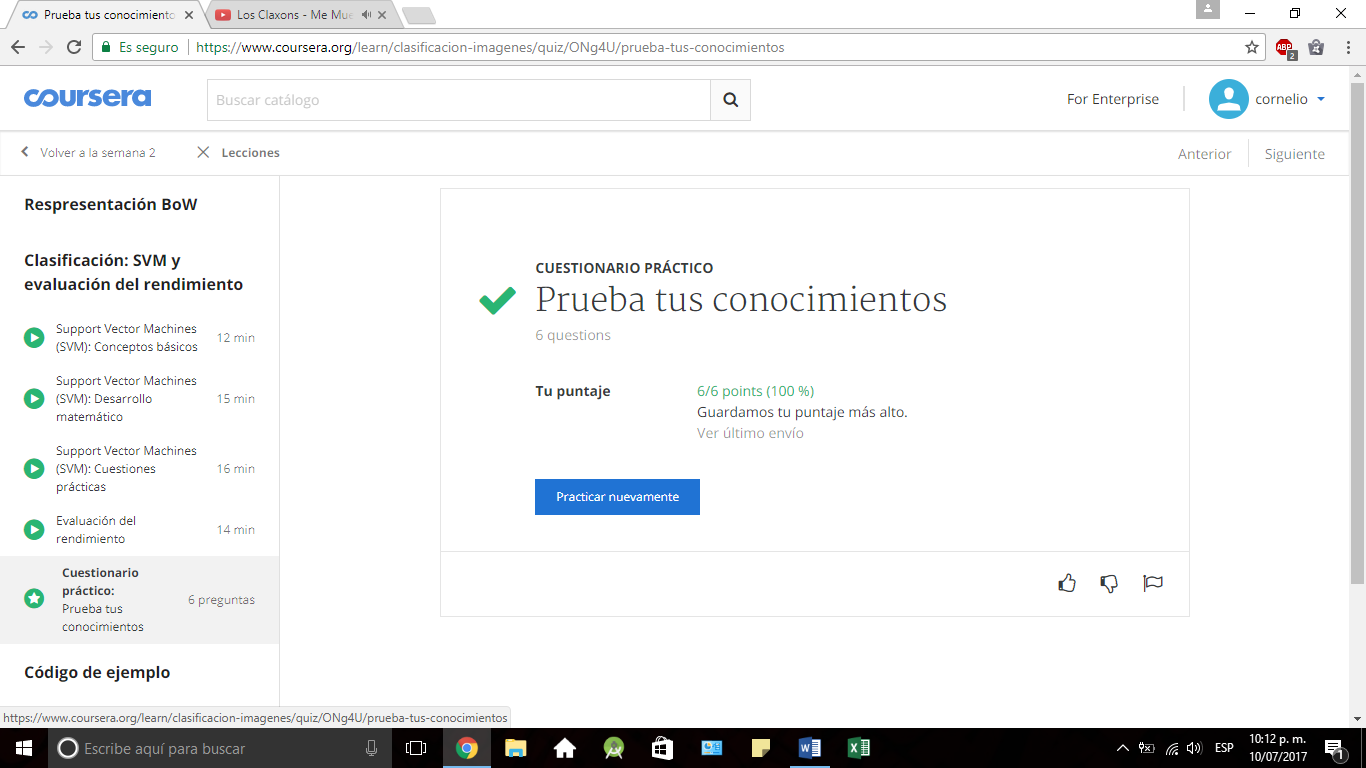


En el siguiente tema de la semana dos se habló de lo que son las máquinas de vectores de soporte, y lo que los compone, como los hiperplanos, los vectores, los vectores de soporte, el margen.

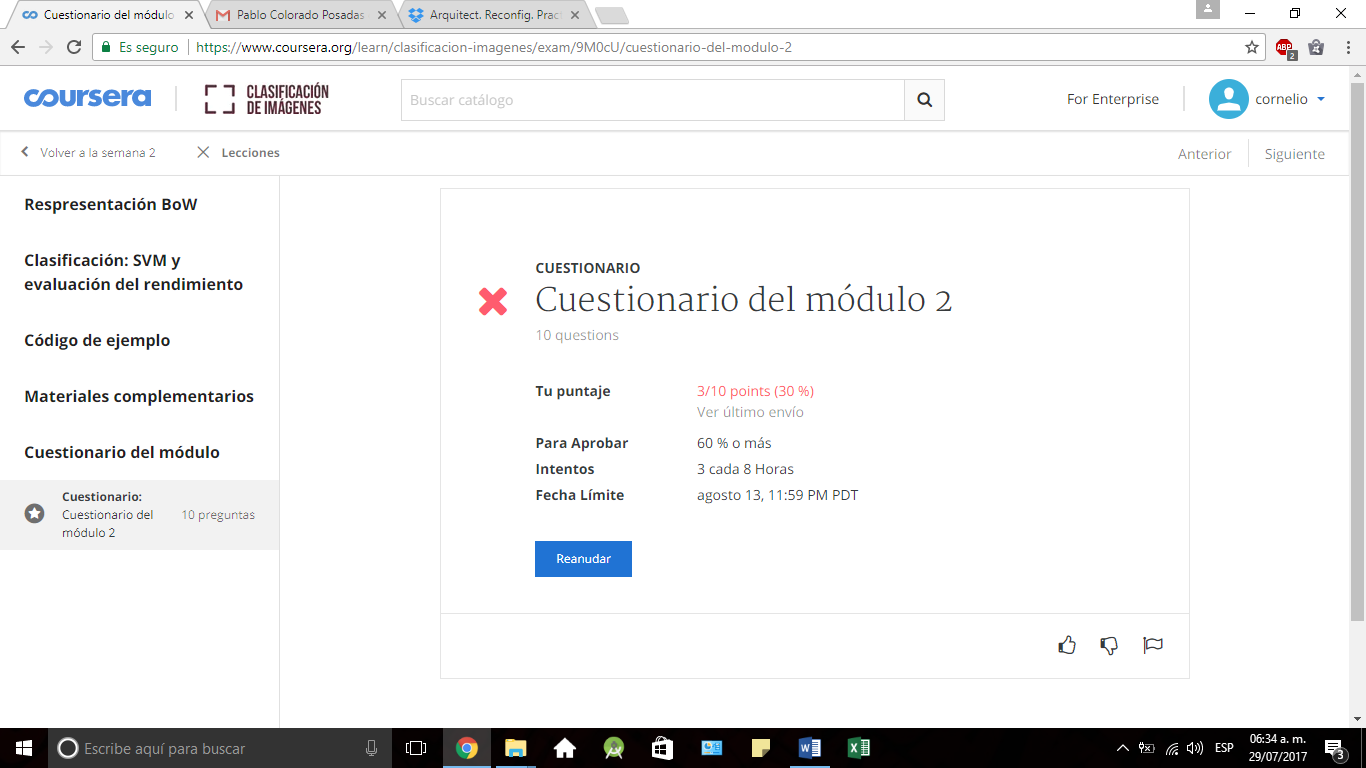
Nos dice que el hiperplano de separación se puede trazar en cualquier parte, por eso hay que buscar la que nos dé un margen mayor ya que será la mejor para la clasificación.



El cuestionario práctico consto de seis preguntas y una vez concluido se pasó a realizar la evaluación final de la semana dos.



En el cuestionario del módulo se respondieron 10 preguntas y en el cual para aprobar se necesitaba el 60% como en cada uno de los cuestionarios.



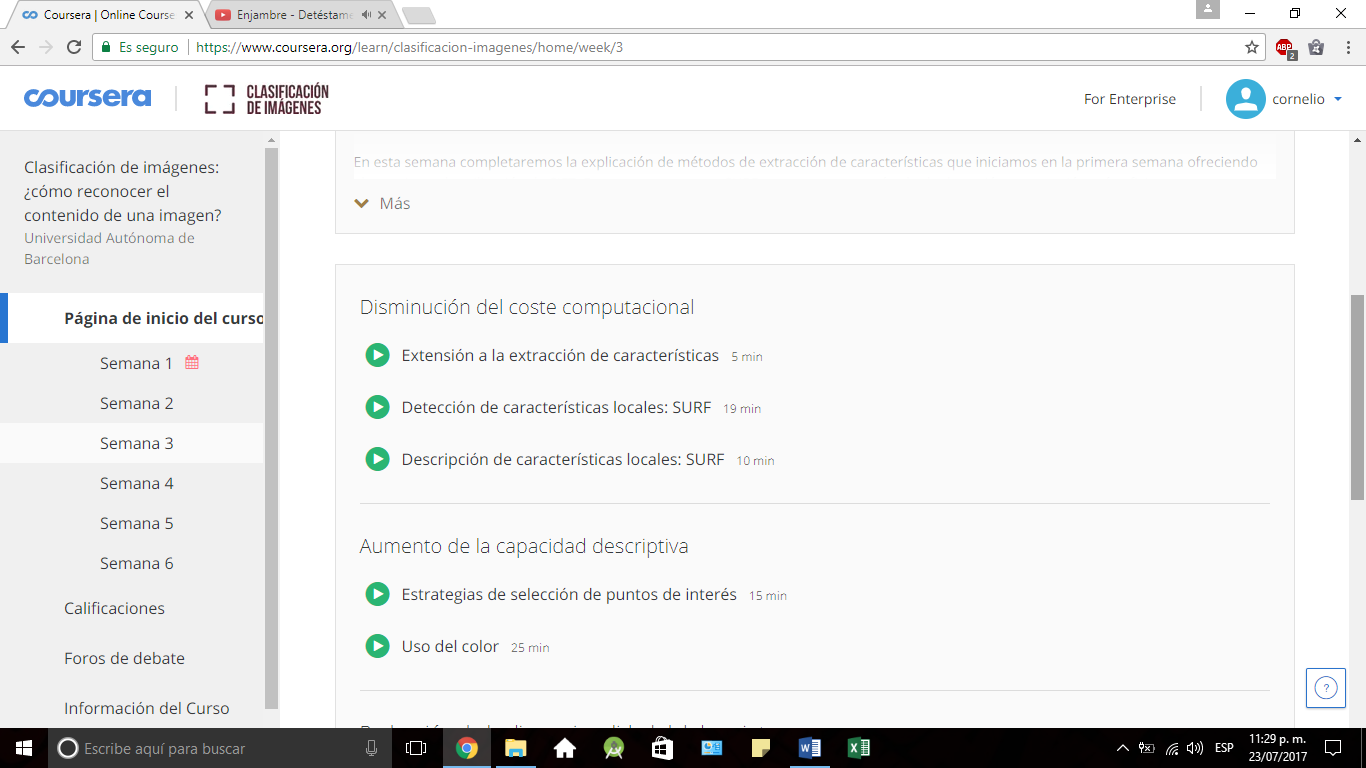
**SEMANA 3**

Para esta semana número tres se basó en la etapa de extracción de características y en la transición a la construcción de un vocabulario, que se vieron al principio.

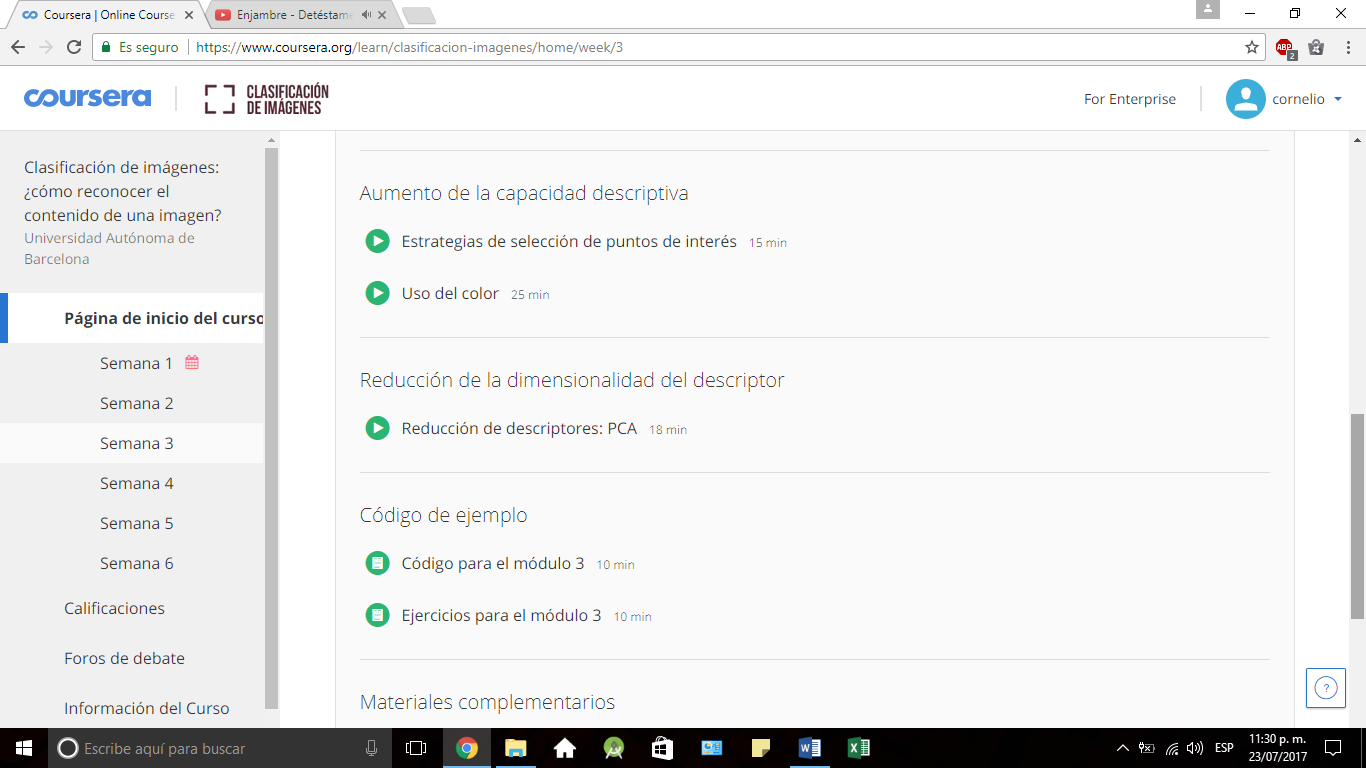
Existen otras alternativas, y entre ellas la de usar los descriptores o usar diferentes tipos de muestreos. Se vieron las ventajas y desventajas de cada uno de ellos. Básicamente el descriptor es algo más compacto, los muestreos permiten cubrir toda la zona de la imagen que en el caso de descriptores no cubren, pero la desventaja evidentemente será el coste computacional y de memoria. Y se buscó cómo solucionar este problema. Se vio como ampliar la capacidad de describir los puntos.

Se aprendió que el color es importante porque si se tiene un tipo de imágenes, como con los que trabaja SIFT, en imágenes en blanco y negro, en grises, nos es imposible distinguir cuál es el color de cada una de las imágenes que se tienen presentes. Sin embargo con color es una tarea extremadamente fácil, y es lo que se propuso, 2 esquemas, 2 técnicas para introducir el color dentro del descriptor. Se vio como conseguir mayor velocidad de respuesta.

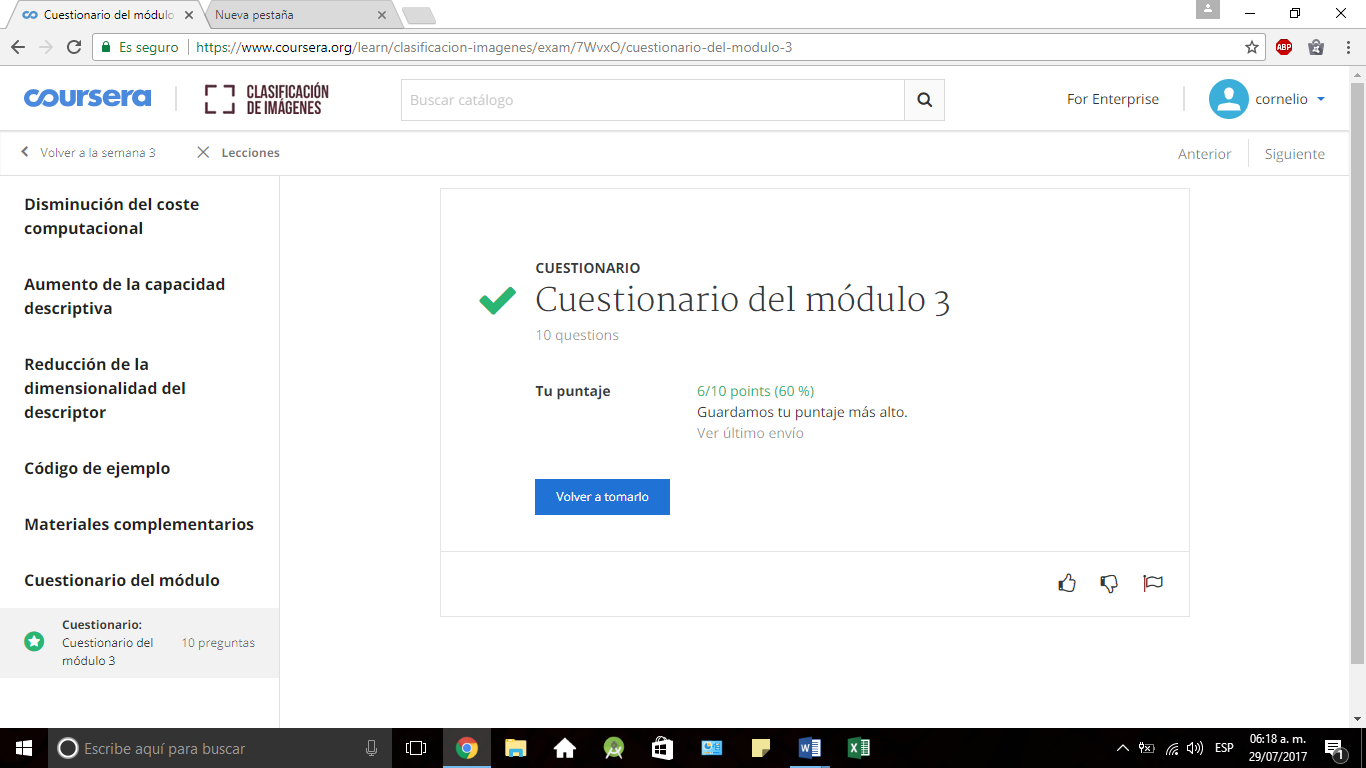
Se vio que SURF detecta tres veces más rápido los puntos de interés a diferencia de SIFT.



 Se vio el muestreo basado en características locales, el problema que comportan, el muestreo denso, y la necesidad de recursos computacionales y hemos visto la solución que era el muestreo aleatorio y cómo esto nos permite obtener unos resultados aceptables sin disparar el coste computacional y de memoria.



En la semana tres no se hicieron evaluaciones de los temas que se vieron y se pasó directo al cuestionario final.



**SEMANA 4.**

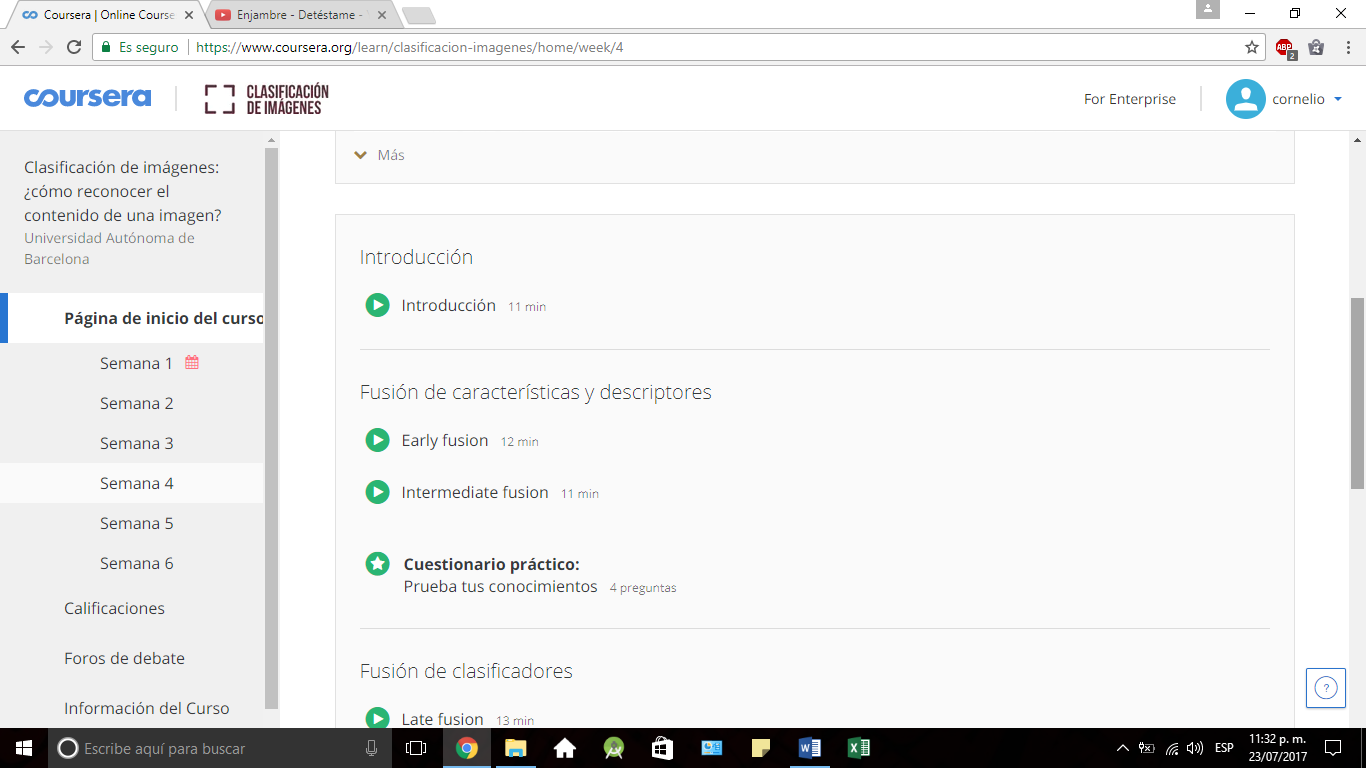
En esta semana se vieron 3 estrategias de fusión.

Se vio como combinar distintos descriptores antes de construir el vocabulario. Esto se llama Early Fusion o fusión temprana. Posteriormente, se explicó cómo combinar distintos vocabularios visuales generados a partir de descriptores de distintos tipos que permiten  identificar regiones comunes entre imágenes de una misma clase, a la vez, permitirá un cierto grado controlado de generalización en la representación. Esta estrategia se acostumbra llamar fusión intermedia o Intermediate Fusion.

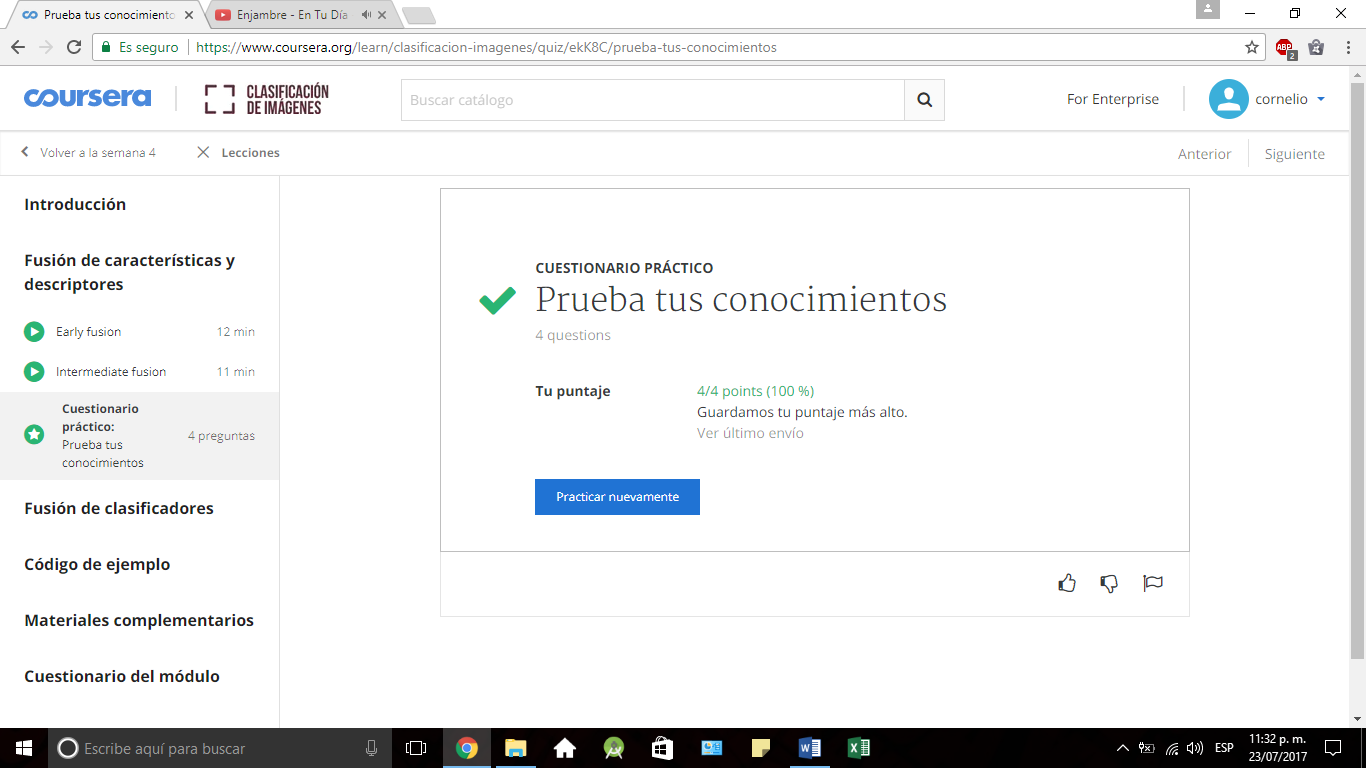
En primer lugar, se ha puesto de relieve que no existe un único descriptor, ni un único vocabulario visual, ni un único clasificador, que sean lo suficientemente potentes como para modelizar la altísima variabilidad que existe en las imágenes correspondientes a determinadas clases visuales.

Por este motivo, se ha justificado que existe la necesidad de combinar distintas fuentes de información para incrementar el rendimiento de la clasificación de imágenes, ya que muchas veces resulta beneficioso tener en cuenta distintos aspectos de una imagen para poder mejorar la representatividad de una clase concreta.

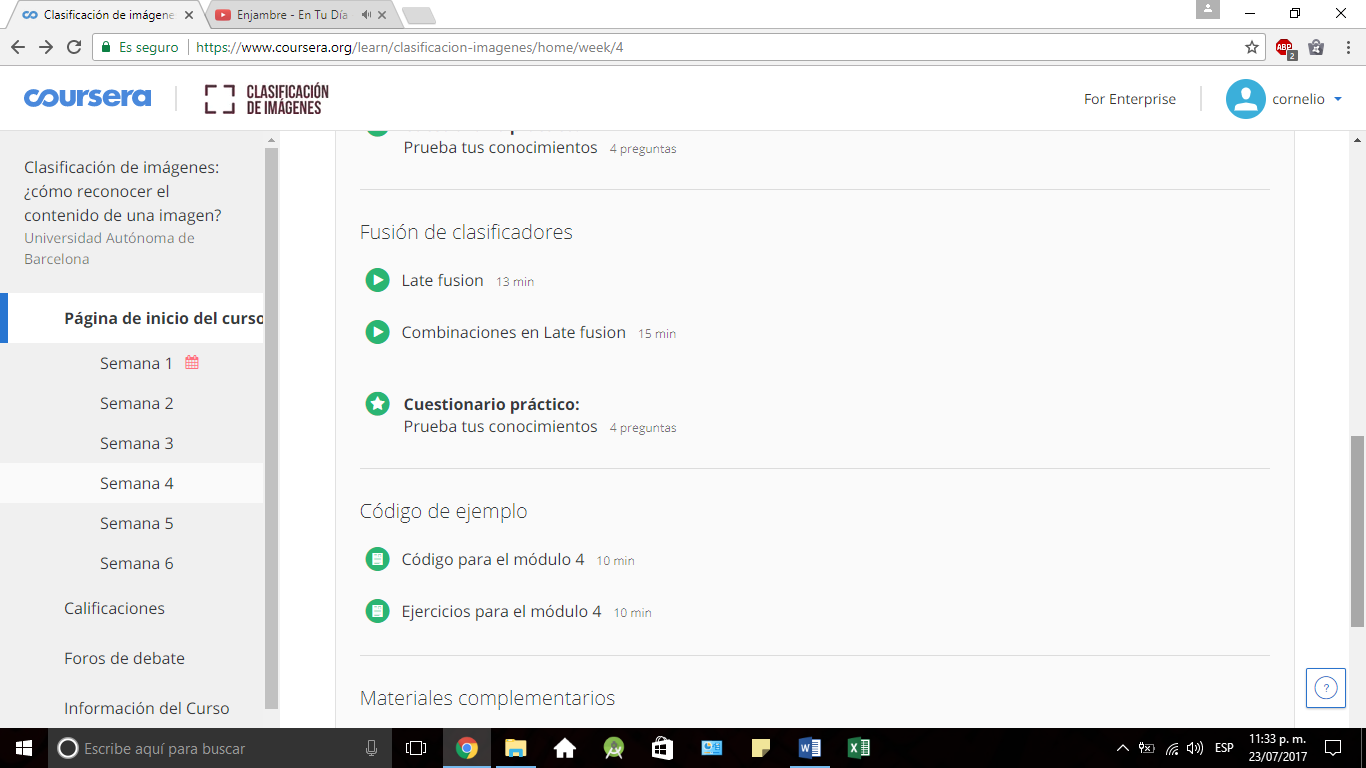
Y finalmente, hemos visto las 3 estrategias básicas de fusión. Será la Early, a nivel de descriptor, Intermediate, a nivel de vocabulario y la Late a nivel de clasificador.



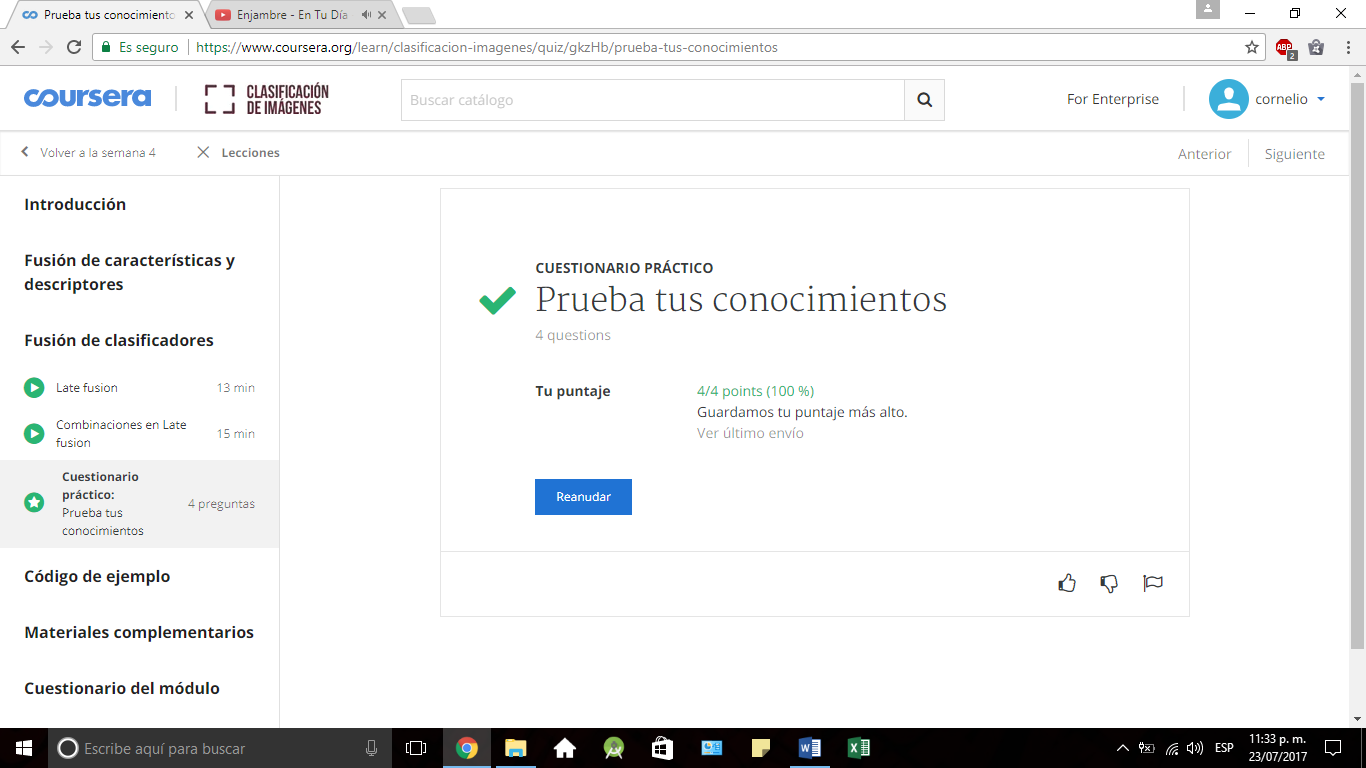
Se realizo el cuestionario practico sobre el tema de fusion de caracteristicas y descriptores, el cual contaba con 4 preguntas.



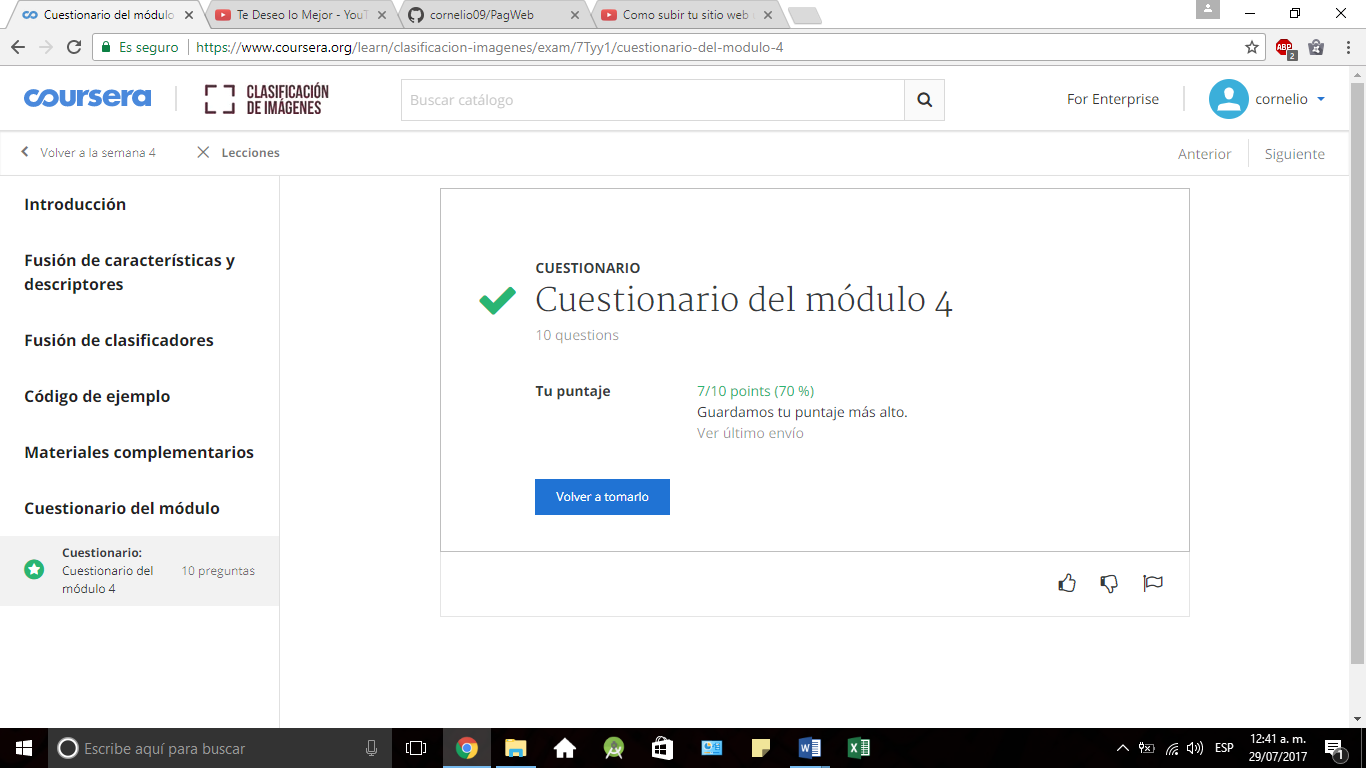
Se vio cómo combinar distintos clasificadores que previamente se habrán especializado en diferentes aspectos de las imágenes de una misma clase. Esta familia de estrategias, llamadas Late Fusion o fusión tardía, nos permitirá que se pueda modelizar una gran variabilidad de contenido visual para una misma clase.



Se realizó el cuestionario del tema, fusión de clasificadores el cual consto de 4 preguntas.



Y por último el cuestionario del módulo de la semana 4, el cual como todos los anteriores conto con 10 preguntas.



**SEMANA 5.**

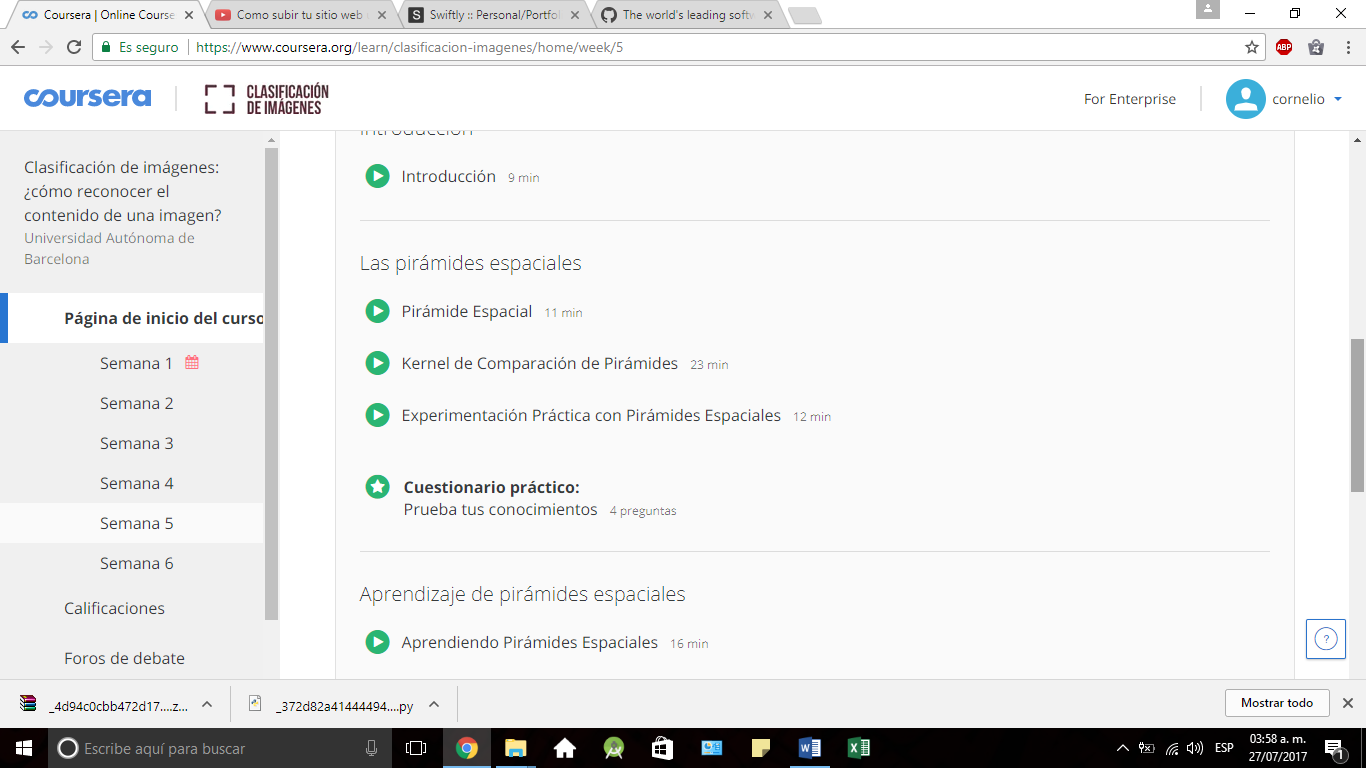
En la semana 5 se habló sobre lo que son las pirámides espaciales, y la cual conto con dos cuestionarios prácticos más aparate el cuestionario final.

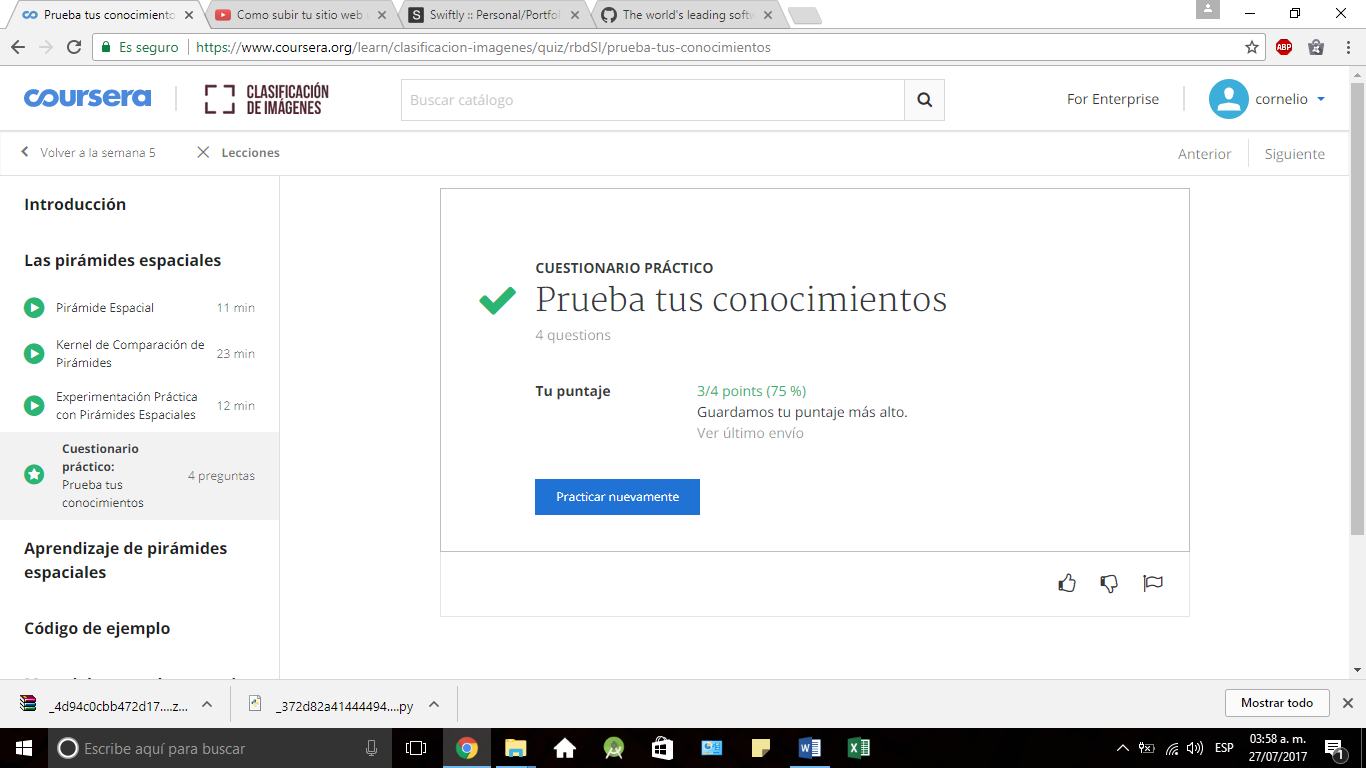
Las pirámides espaciales consisten en colocar unas rejillas cada vez más finas en una imagen y en combinar las palabras visuales de cada rejilla y a diferentes niveles de resolución.

Se ha visto cómo la incorporación de información espacial con pirámides es básicamente una concatenación ponderada de histogramas de palabras extraídas en distintas regiones y ponderadas a distintos niveles de resolución.  Es decir, el histograma de cada región y nivel representa el número de veces que aparece cada palabra del vocabulario visual en esa región en concreto. Pero la magnitud de los valores de los histogramas en niveles muy bajos, como es menor en niveles altos a considerar porciones muy pequeñas de la imagen, tendrán que ser dados más prioridad o más importancia para el cálculo del histograma concatenado final.

El kernel de comparación su objetivo es el cálculo de correspondencia ante descriptores visuales de dos imágenes en distinta resolución, para poder estimar su similitud de una forma muy eficiente.

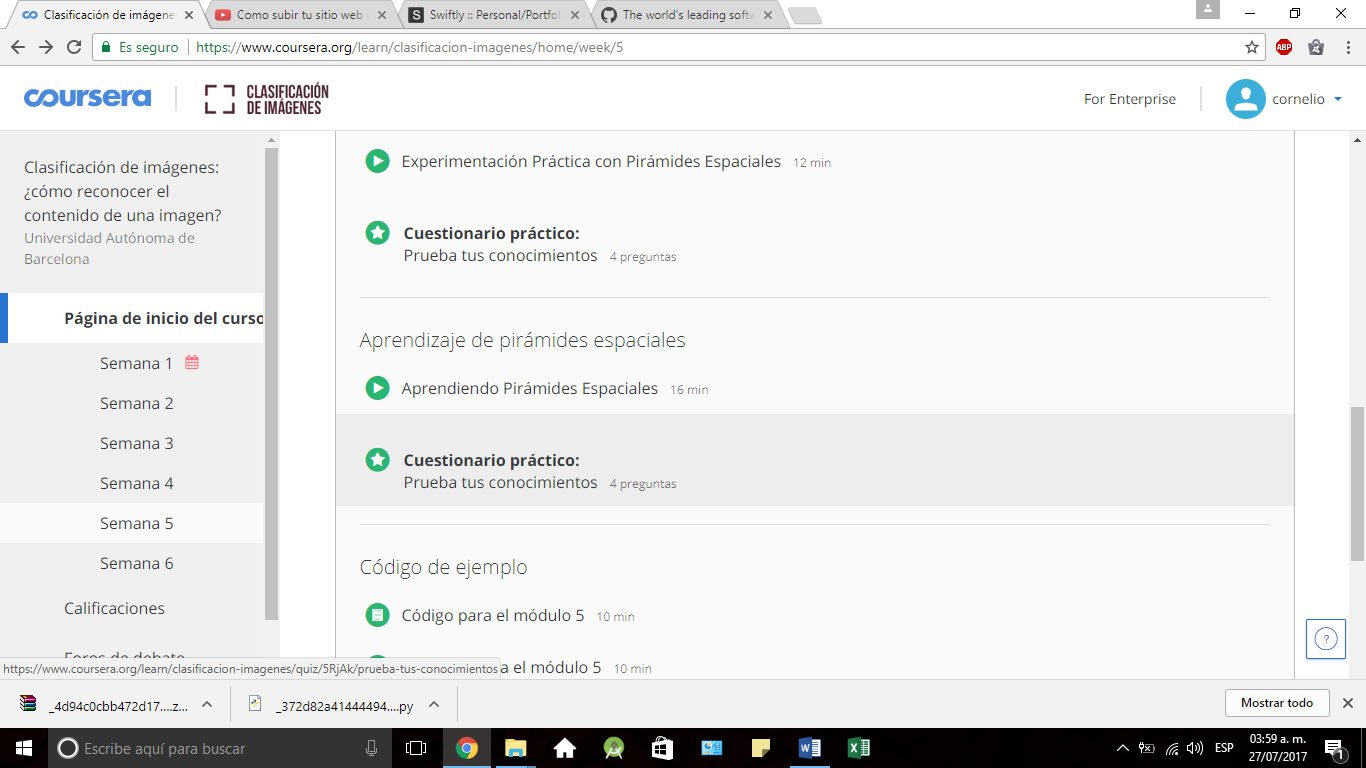
La idea principal es concatenar histogramas de varios niveles, donde el histograma tiene en cada nivel el doble de bins o dimensiones que el del nivel posterior.

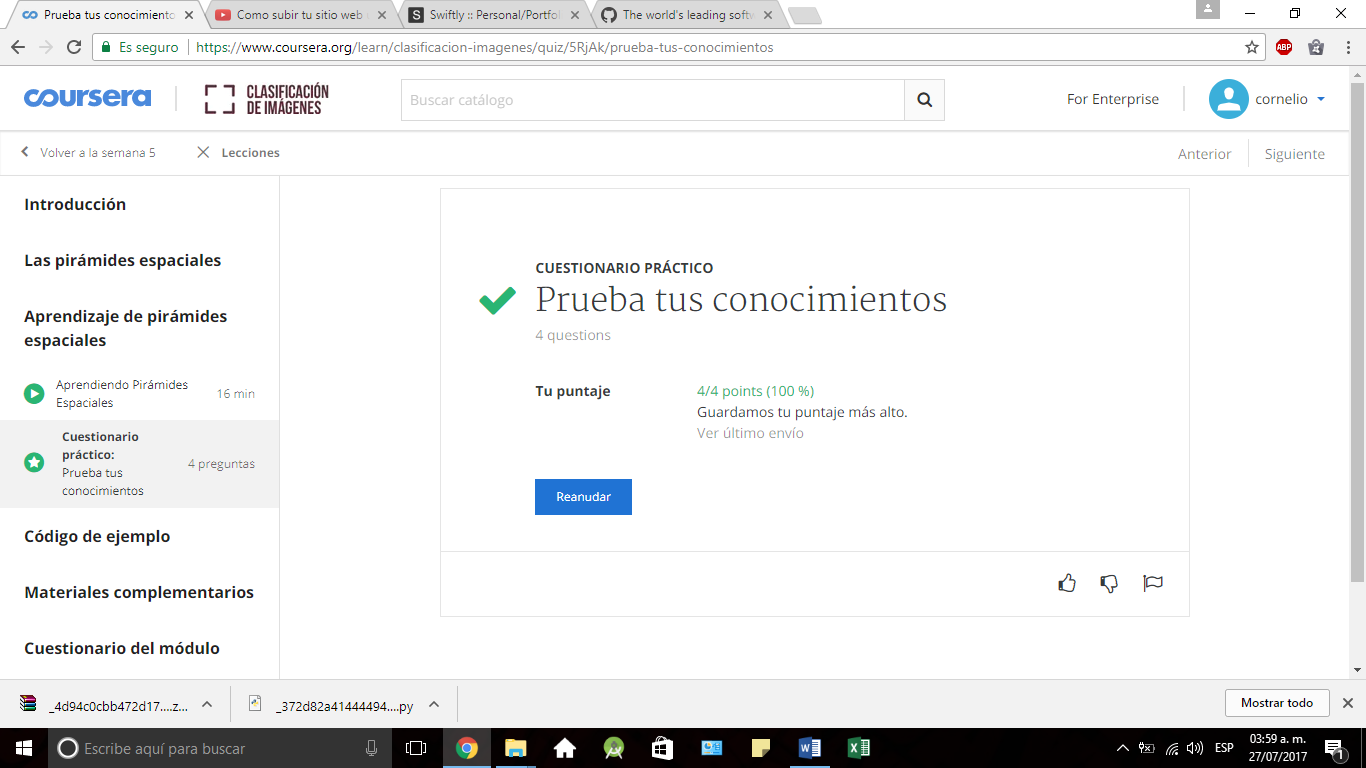


Se realizó el primer cuestionario práctico de la semana 5 el cual consto de 4 preguntas.

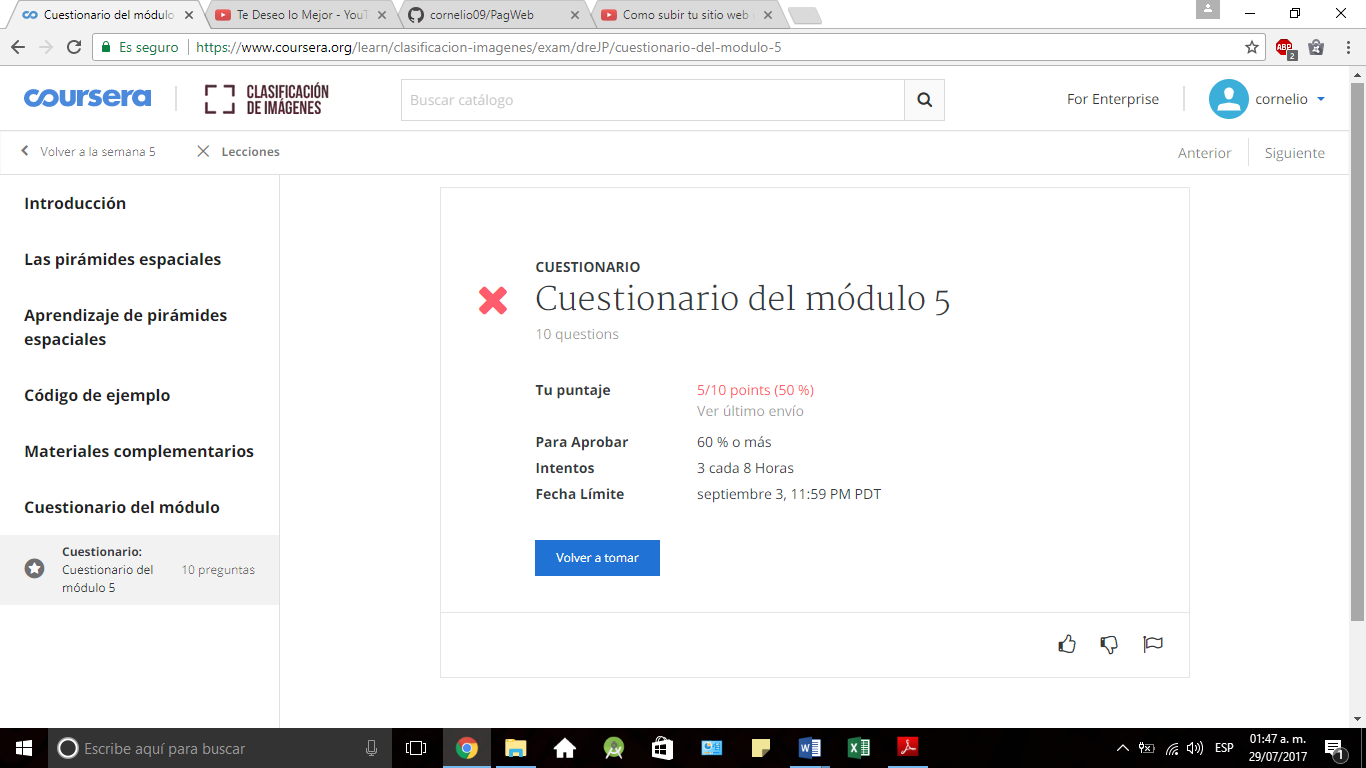
En el video aprendiendo pirámides espaciales, se ha enfatizado que el uso de mallas 2x2 en el esquema original de las pirámides espaciales en el Bag of Words pueden no ser la más apropiadas para las categorías de imágenes.

Así, se ha presentado cómo utilizar la teoría A y B como una metodología de fusión de muestras a partir de la medida de información mutua que existe entre muestras y etiquetas.



Se realizó el cuestionario practico numero dos el cual también conto con 4 preguntas.

Y por último el cuestionario final de la semana 5.



**SEMANA 6.**

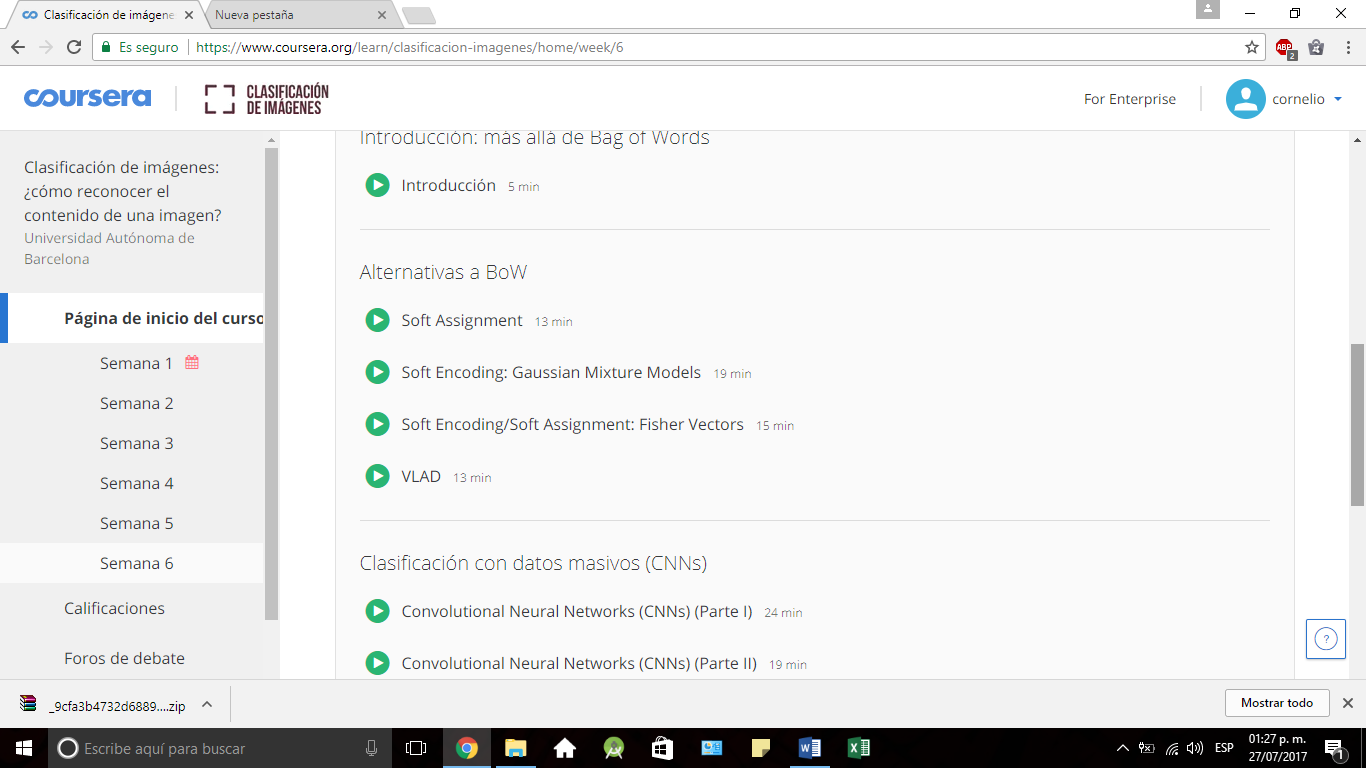
En la semana 6 se habló de algunas alternativas a bag of words pero que son relevantes para el proceso de clasificación de imágenes, en esta última semana para la evaluación solo hubo el examen final de la semana 6 y un examen final de todo el curso.

Hasta el momento lo que se tenía era la estructura de clasificación. Se tenía una fase de extracción de características, con detección, descripción, una fase de representación, una fase de pirámides espaciales y una fase final de clasificación. Se conoce como esquema de bag-of-words.

Se vio lo que se conoce como soft assignment. Y que es una evolución dentro del esquema de bag of words, para la representación con histogramas de palabras para una imagen dada.

El objetivo es tomar un conjunto grande y variable de puntos de interés de una imagen de entrada y transformar esta información en un único vector más compacto y que explique el contenido de la imagen.

Se habló de Soft encoding, en el cual se utiliza k-means y se determinan un conjunto de centros del espacio n dimensional que son los centros de las palabras. También se vio que utiliza  Modelos de mezcla gaussianas para resolver este tipo de problemas de hard encoding.



Se llegó hasta este punto ya que por razones de tiempo no se han realizado las ultimas evaluaciones las cuales se harán pero ya no se pudo capturar y de igual manera se tratará de aprobar los cuestionarios finales de cada semana ya que solo se han aprobado dos semanas.